```
                         GenCore version 4.5
                Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 04:22:13 ; Search time 111.26 Seconds
                                             (without alignments)
                                             7.553 Million cell updates/sec

Title:          US-09-529-121-10
Perfect score:  53
Sequence:       1 YLHDPEFNL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters:      436

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :      SPTREMBL_14:*
               1:  sp_archea:*
               2:  sp_bacteria:*
               3:  sp_fungi:*
               4:  sp_human:*
               5:  sp_invertebrate:*
               6:  sp_mammal:*
               7:  sp_mhc:*
               8:  sp_organelle:*
               9:  sp_phage:*
               10:  sp_plant:*
               11:  sp_rodent:*
               12:  sp_virus:*
               13:  sp_vertebrate:*
               14:  sp_unclassified:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                 %
Result          Query
  No.   Score   Match Length DB  ID                  Description
-------------------------------------------------------------------------
     1     20    37.7      7  6  Q28742                  Q28742 oryctolagus
     2     18    34.0      7 13   P82101                  P82101 litoria rub
     3     17    32.1      8  4  Q15889                  Q15889 homo sapien
     4     16    30.2      8  4  Q16468                  Q16468 homo sapien
     5     15    28.3      8  4  Q15888                  Q15888 homo sapien
     6     15    28.3      8 11   Q9QVJ8                  Q9qvj8 mus sp. mep
     7     15    28.3      8 12   Q9WJ33                  Q9wj33 pseudorabie
     8     14    26.4      9  2  Q47556                  Q47556 escherichia
     9     14    26.4      9  5  O96417                  O96417 drosophila
    10     14    26.4      9  8  O78337                  O78337 caloglossa
    11     14    26.4      9  8  Q9TLD0                  Q9tld0 bostrychia
    12     14    26.4      9  8  Q9TJ87                  Q9tj87 caloglossa
    13     14    26.4      9  8  Q9TJ85                  Q9tj85 caloglossa
    14     14    26.4      9  8  Q9T389                  Q9t389 caloglossa
    15     14    26.4      9  8  Q9T388                  Q9t388 caloglossa
    16     14    26.4      9  8  Q9T387                  Q9t387 bostrychia
    17     14    26.4      9 10   O81962                  O81962 caloglossa
    18     14    26.4      9 10   O81964                  O81964 caloglossa
    19     14    26.4      9 10   O81966                  O81966 caloglossa
    20     14    26.4      9 10   O81968                  O81968 caloglossa
    21     14    26.4      9 10   O82778                  O82778 caloglossa
    22     13    24.5      7  8  O99182                  O99182 pterolebias
    23     13    24.5      8  2  Q9R7T2                  Q9r7t2 escherichia
    24     13    24.5      8  4  Q9UMH9                  Q9umh9 homo sapien
    25     13    24.5      8  6  Q9TT78                  Q9tt78 canis famil
    26     13    24.5      9  2  Q51765                  Q51765 pseudomonas
    27     13    24.5      9  2  Q9R7H9                  Q9r7h9 haemophilus
    28     13    24.5      9  2  Q9R735                  Q9r735 streptomyce
    29     13    24.5      9  4  P78484                  P78484 homo sapien
    30     13    24.5      9  4  Q9UM87                  Q9um87 homo sapien
    31     13    24.5      9  4  Q9UC36                  Q9uc36 homo sapien
    32     13    24.5      9  5  Q9TWD6                  Q9twd6 leptinotars
    33     13    24.5      9  9  Q9XJN0                  Q9xjn0 bacteriopha
    34     13    24.5      9 12   Q85723                  Q85723 simian sarc
    35     12    22.6      5 13   P82099                  P82099 litoria rub
    36     12    22.6      7 12   Q66205                  Q66205 porcine tra
    37     12    22.6      8  2  Q9R5L7                  Q9r517 clostridium
    38     12    22.6      8  3  P87225                  P87225 saccharomyc
    39     12    22.6      8  3  O13591                  O13591 saccharomyc
    40     12    22.6      8  5  Q9VPK2                  Q9vpk2 drosophila
    41     12    22.6      8  8  Q9XNP8                  Q9xnp8 boophilus m
    42     12    22.6      8  9  Q37854                  Q37854 bacteriopha
    43     12    22.6      8 11   Q9QVI5                  Q9qvi5 rattus sp.
    44     12    22.6      9  2  Q43960                  Q43960 azotobacter
    45     12    22.6      9  4  Q16605                  Q16605 homo sapien
    46     12    22.6      9  4  Q14715                  Q14715 homo sapien
    47     12    22.6      9  4  Q9UCQ9                  Q9ucq9 homo sapien
    48     12    22.6      9 10   P82440                  P82440 nicotiana t
    49     12    22.6      9 11   Q61723                  Q61723 mus musculu
    50     12    22.6      9 11   Q9QVH9                  Q9qvh9 mus sp. sup
    51     12    22.6      9 12   Q66545                  Q66545 human herpe
    52     11    20.8      5 13   P82071                  P82071 litoria rub
    53     11    20.8      5 13   P82072                  P82072 litoria rub
    54     11    20.8      7  2  O07354                  O07354 synechococc
    55     11    20.8      7  5  Q9VYN9                  Q9vyn9 drosophila
    56     11    20.8      7 12   Q9YQ10                  Q9yq10 porcine tra
    57     11    20.8      7 13   P82065                  P82065 litoria rub
    58     11    20.8      8  2  Q9RQ57                  Q9rq57 buchnera ap
    59     11    20.8      8  2  Q9RQ49                  Q9rq49 buchnera ap
    60     11    20.8      8  2  Q9R3X0                  Q9r3x0 planktothri
    61     11    20.8      8  3  Q05403                  Q05403 saccharomyc
    62     11    20.8      8  4  Q15901                  Q15901 homo sapien
    63     11    20.8      8  4  Q9Y4J3                  Q9y4j3 homo sapien
    64     11    20.8      8  4  Q9UL56                  Q9ul56 homo sapien
    65     11    20.8      8  4  Q9UDZ4                  Q9udz4 homo sapien
    66     11    20.8      8  6  O02831                  O02831 oryctolagus
    67     11    20.8      8  6  Q9XSY1                  Q9xsy1 canis famil
    68     11    20.8      8  8  O19957                  O19957 gossypium h
    69     11    20.8      8  8  Q35792                  Q35792 saccharomyc
    70     11    20.8      8 12   Q83332                  Q83332 murine hepa
    71     11    20.8      8 12   Q84156                  Q84156 orf virus.
    72     11    20.8      9  2  Q57328                  Q57328 aeromonas s
    73     11    20.8      9  2  Q44001                  Q44001 aeromonas e
    74     11    20.8      9  2  Q44377                  Q44377 aeromonas t
    75     11    20.8      9  2  Q44468                  Q44468 aeromonas v

                                 ALIGNMENTS

RESULT   1
Q28742
ID   Q28742      PRELIMINARY;      PRT;     7 AA.
AC   Q28742;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ALPHA-MYOSIN HEAVY CHAIN (FRAGMENT).
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84221901.
RA   Friedman D.J., Umeda P.K., Sinha A.M., Hsu H.J., Jokovcic S.,
RA   Rabinowitz M.;
RT   "Characterization of genomic clones specifying rabbit alpha- and beta-
RT   ventricular myosin heavy chains.";
RL   Proc. Natl. Acad. Sci. U.S.A. 81:3044-3048(1984).
DR   EMBL; K01698; AAA31415.1; -.
KW   Myosin.
FT   NON_TER       1      1
SQ   SEQUENCE   7 AA;  916 MW;  6B1B1AA1E69326B0 CRC64;

  Query Match              37.7%;  Score 20;  DB 6;  Length 7;
  Best Local Similarity   60.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LHDPE 6
              :|| |
Db          3 MHDEE 7


RESULT   2
P82101
ID   P82101      PRELIMINARY;      PRT;     7 AA.
AC   P82101;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ELECTRIN 5.
OS   Litoria rubella (Desert tree frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Hylidae;
OC   Litoria.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=SKIN SECRETION;
RA   Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;
RT   "Peptides from the skin glands of the Australian buzzing tree frog
RT   Litori electrica. Comparison with the skin peptides from Litoria
RT   rubella.";
RL   Aust. J. Chem. 52:0-0(1999).
KW   Amphibian skin; Amidation.
FT   MOD_RES       7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  834 MW;  6DD05B076B0B5030 CRC64;

  Query Match              34.0%;  Score 18;  DB 13;  Length 7;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LHDPE 6
              :::||
Db          1 IYEPE 5


RESULT   3
Q15889
ID   Q15889      PRELIMINARY;      PRT;     8 AA.
AC   Q15889;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   (CLONE XP15H8B) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   Lee C.C., Yazdani A., Wehnert M., Bailey J., Couch L., Xiong M.,
RA   Coolbaugh M.I., Chinault C.A., Baldini A., Lindsay E.A., Zhao Z.Y.,
RA   Caskey C.T.H.;
RL   Hum. Mol. Genet. 0:0-0(0).
DR   EMBL; L32070; AAA73879.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  865 MW;  0474472325A761E7 CRC64;

  Query Match              32.1%;  Score 17;  DB 4;  Length 8;
  Best Local Similarity   42.9%;  Pred. No. 3e+05;
  Matches    3;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 LHDPEFN 8
              ||  : |
Db          1 LHPSKLN 7


RESULT   4
Q16468
ID   Q16468      PRELIMINARY;      PRT;     8 AA.
AC   Q16468;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   DNA FOR COSMID CC13-1134 PCR PRIMER 1 (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96435920.
RA   James L.A., Ogilvie D.J., Yamakawa K., Nakamura Y., Stirling C.J.,
RA   Anand R.;
RT   "Walking, cloning, and mapping with YACs in 3q27: localization of five
RT   ESTs including three members of the cystatin gene family and
RT   identification of CpG islands.";
RL   Genomics 32:425-430(1996).
DR   EMBL; X88976; CAA61407.1; -.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  925 MW;  FD5411A7376871E6 CRC64;

  Query Match              30.2%;  Score 16;  DB 4;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LHDP 5
              :| |
Db          1 MHGP 4


RESULT   5
Q15888
ID   Q15888      PRELIMINARY;      PRT;     8 AA.
AC   Q15888;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   (CLONE XP15H8A) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   Lee C.C., Yazdani A., Wehnert M., Bailey J., Couch L., Xiong M.,
RA   Coolbaugh M.I., Chinault C.A., Baldini A., Lindsay E.A., Zhao Z.Y.,
RA   Caskey C.T.H.;
RL   Hum. Mol. Genet. 0:0-0(0).
DR   EMBL; L32069; AAA73878.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  1068 MW;  0315A37EAB5B0763 CRC64;
```

```
  Query Match             28.3%;  Score 15;  DB 4;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 PEF 7
            ||:
Db        2 PEY 4

RESULT   6
Q9QVJ8
ID   Q9QVJ8      PRELIMINARY;      PRT;     8 AA.
AC   Q9QVJ8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MEPRIN-A.
OS   Mus sp.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 91373354.
RA   Kounnas M.Z., Wolz R.L., Gorbea C.M., Bond J.S.;
RT   "Meprin-A and -B. Cell surface endopeptidases of the mouse kidney.";
RL   J. Biol. Chem. 266:17350-17357(1991).
SQ   SEQUENCE   8 AA;  877 MW;  43A5A76AB4069DD4 CRC64;


  Query Match             28.3%;  Score 15;  DB 11;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        2 LHDP 5
            : ||
Db        3 MRDP 6

RESULT   7
Q9WJ33
ID   Q9WJ33      PRELIMINARY;      PRT;     8 AA.
AC   Q9WJ33;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   SUID HERPESVIRUS 1 PUTATIVE UL47 AND UL46 GENES AND PARTIAL GB
DE   GENE, STRAIN KAPLAN (FRAGMENT).
GN   GB.
OS   Pseudorabies virus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KAPLAN;
RA   Bras F., Dezelee S., Simonet B., Nguyen X., Vende P., Flamand A.,
RA   Masse M.J.;
RT   "The left border of the genomic inversion of pseudorabies virus
RT   contains genes homologous to the UL46 and UL47 genes of Herpes Simplex
RT   Virus type 1, but no UL45 Gene.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ010303; CAA09075.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  875 MW;  262DDAB76AAB05BB CRC64;


  Query Match             28.3%;  Score 15;  DB 12;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        4 DPE 6
            ||:
Db        4 DPD 6
```

```
RESULT   8
Q47556
ID   Q47556      PRELIMINARY;      PRT;     9 AA.
AC   Q47556;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ASPARTATE TRANSCARBAMOYLASE REGULATORY CHAIN (FRAGMENT).
GN   PYRI.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 82275057.
RA   Pauza C.D., Karels M.J., Navre M., Schachman H.K.;
RT   "Genes encoding Escherichia coli aspartate transcarbamoylase: the
RT   pyrB-pyrI operon.";
RL   Proc. Natl. Acad. Sci. U.S.A. 79:4020-4024(1982).
RN   [2]
RP   SEQUENCE OF 1-5 FROM N.A.
RX   MEDLINE; 83195078.
RA   Hoover T.A., Roof W.D., Foltermann K.F., O'Donovan G.A., Bencini D.A.,
RA   Wild J.R.;
RT   "Nucleotide sequence of the structural gene (pyrB) that encodes the
RT   catalytic polypeptide of aspartate transcarbamoylase of Escherichia
RT   coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 80:2462-2466(1983).
DR   EMBL; J01670; AAA24475.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1085 MW;  99EFD723344AA1F1 CRC64;


  Query Match             26.4%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 HD 4
            ||
Db        3 HD 4


RESULT   9
O96417
ID   O96417      PRELIMINARY;      PRT;     9 AA.
AC   O96417;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   SXL E1 FORM (FRAGMENT).
GN   SXL.
OS   Drosophila virilis (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98337843.
RA   Erickson J.W., Cline T.W.;
RT   "Key aspects of the primary sex determination mechanism are conserved
RT   across the genus Drosophila.";
RL   Development 125:3259-3268(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Bell M., Cline T.W.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF046045; AAC97605.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1089 MW;  9A8BD1AAA9C449CA CRC64;
```

```
  Query Match             26.4%;  Score 14;  DB 5;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 EFN 8
           :||
Db       2 DFN 4

RESULT  10
O78337
ID   O78337      PRELIMINARY;      PRT;     9 AA.
AC   O78337;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa leprieurii.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=736;
RA   Kamiya M., West J.A., King R.J., Zuccarello G.C., Tanaka J., Hara Y.;
RT   "Evolutionary divergence in the red algae Caloglossa leprieurii and C.
RT   apomeiotica.";
RL   J. Phycol. 34:361-370(1998).
DR   EMBL; D87813; BAA31279.1; -.
KW   Chloroplast.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             26.4%;  Score 14;  DB 8;  Length 9;
  Best Local Similarity   22.2%;  Pred. No. 3e+05;
  Matches    2;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 YLHDPEFNL 9
           ::  |  |:
Db       1 FVETPTANV 9

RESULT  11
Q9TLD0
ID   Q9TLD0      PRELIMINARY;      PRT;     9 AA.
AC   Q9TLD0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Bostrychia moritziana.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC   Bostrychia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M3189.CO;
RA   Zuccarello G.C., West J.A., King R.J.;
RT   "Biogeography of Bostrychia moritziana (Ceramiales).";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF126705; AAD55863.1; -.
KW   Chloroplast.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             26.4%;  Score 14;  DB 8;  Length 9;
  Best Local Similarity   22.2%;  Pred. No. 3e+05;
```

```
  Matches    2;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 YLHDPEFNL 9
           ::  |  |:
Db       1 FVETPTANV 9

RESULT  12
Q9TJ87
ID   Q9TJ87      PRELIMINARY;      PRT;     9 AA.
AC   Q9TJ87;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa monosticha.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=987;
RA   Kamiya M., Tanaka J., King R.J., West J.A., Zuccarello G.C., Kawai H.;
RT   "Reproductive and genetic distinction between broad and narrow
RT   entities of Caloglossa continua (Delesseriaceae, Rhodophyta).";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB023381; BAA88912.1; -.
KW   Chloroplast.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             26.4%;  Score 14;  DB 8;  Length 9;
  Best Local Similarity   22.2%;  Pred. No. 3e+05;
  Matches    2;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 YLHDPEFNL 9
           ::  |  |:
Db       1 FVETPTANV 9

RESULT  13
Q9TJ85
ID   Q9TJ85      PRELIMINARY;      PRT;     9 AA.
AC   Q9TJ85;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa stipitata.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Kamiya M., Tanaka J., King R.J., West J.A., Zuccarello G.C., Kawai H.;
RT   "Reproductive and genetic distinction between broad and narrow
RT   entities of Caloglossa continua (Delesseriaceae, Rhodophyta).";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB023384; BAA88918.1; -.
KW   Chloroplast.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             26.4%;  Score 14;  DB 8;  Length 9;
  Best Local Similarity   22.2%;  Pred. No. 3e+05;
  Matches    2;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy       1 YLHDPEFNL 9
           ::  |  |:
Db       1 FVETPTANV 9


RESULT  14
Q9T389
ID   Q9T389      PRELIMINARY;      PRT;     9 AA.
AC   Q9T389;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa postiae.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAW1122, AND 962;
RA   Kamiya M., Tanaka J., King R.J., West J.A., Zuccarello G.C., Kawai H.;
RT   "Reproductive and genetic distinction between broad and narrow
RT   entities of Caloglossa continua (Delesseriaceae, Rhodophyta).";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB023383; BAA88916.1; -.
DR   EMBL; AB023382; BAA88914.1; -.
KW   Chloroplast.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;


  Query Match             26.4%;  Score 14;  DB 8;  Length 9;
  Best Local Similarity   22.2%;  Pred. No. 3e+05;
  Matches    2;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 YLHDPEFNL 9
           ::  |  |:
Db       1 FVETPTANV 9


RESULT  15
Q9T388
ID   Q9T388      PRELIMINARY;      PRT;     9 AA.
AC   Q9T388;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa continua.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=729, AND 500;
RA   Kamiya M., Tanaka J., King R.J., West J.A., Zuccarello G.C., Kawai H.;
RT   "Reproductive and genetic distinction between broad and narrow
RT   entities of Caloglossa continua (Delesseriaceae, Rhodophyta).";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB023380; BAA88910.1; -.
DR   EMBL; AB023379; BAA88908.1; -.
KW   Chloroplast.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;


  Query Match             26.4%; .Score 14;  DB 8;  Length 9;
  Best Local Similarity   22.2%;  Pred. No. 3e+05;
```

```
  Matches    2;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 YLHDPEFNL 9
           ::  |  |:
Db       1 FVETPTANV 9


RESULT  16
Q9T387
ID   Q9T387      PRELIMINARY;      PRT;     9 AA.
AC   Q9T387;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Bostrychia radicans.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC   Bostrychia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R.3881.QLD, M3001.MI, R3826.BZ, AND B.R.F.M.BZ;
RA   Zuccarello G.C., West J.A., King R.J.;
RT   "Biogeography of Bostrychia moritziana (Ceramiales).";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF126715; AAD55883.1; -.
DR   EMBL; AF126701; AAD55855.1; -.
DR   EMBL; AF126704; AAD55861.1; -.
DR   EMBL; AF126706; AAD55865.1; -.
KW   Chloroplast.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;


  Query Match             26.4%;  Score 14;  DB 8;  Length 9;
  Best Local Similarity   22.2%;  Pred. No. 3e+05;
  Matches    2;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 YLHDPEFNL 9
           ::  |  |:
Db       1 FVETPTANV 9


RESULT  17
O81962
ID   O81962      PRELIMINARY;      PRT;     9 AA.
AC   O81962;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa apomeiotica.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=910;
RA   Kamiya M., West J.A., King R.J., Zuccarello G.C., Tanaka J., Hara Y.;
RT   "Evolutionary divergence in the red algae Caloglossa leprieurii and C.
RT   apomeiotica.";
RL   J. Phycol. 34:361-370(1998).
DR   EMBL; D89948; BAA31281.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;


  Query Match             26.4%;  Score 14;  DB 10;  Length 9;
  Best Local Similarity   22.2%;  Pred. No. 3e+05;
  Matches    2;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy       1 YLHDPEFNL 9
           :: |  |:
Db       1 FVETPTANV 9


RESULT  18
O81964
ID   O81964      PRELIMINARY;      PRT;     9 AA.
AC   O81964;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa continua.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=639;
RA   Kamiya M., West J.A., King R.J., Zuccarello G.C., Tanaka J., Hara Y.;
RT   "Evolutionary divergence in the red algae Caloglossa leprieurii and C.
RT   apomeiotica.";
RL   J. Phycol. 34:361-370(1998).
DR   EMBL; D89950; BAA31285.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             26.4%;  Score 14;  DB 10;  Length 9;
  Best Local Similarity   22.2%;  Pred. No. 3e+05;
  Matches    2;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 YLHDPEFNL 9
           :: |  |:
Db       1 FVETPTANV 9


RESULT  19
O81966
ID   O81966      PRELIMINARY;      PRT;     9 AA.
AC   O81966;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa monosticha.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=892;
RA   Kamiya M., West J.A., King R.J., Zuccarello G.C., Tanaka J., Hara Y.;
RT   "Evolutionary divergence in the red algae Caloglossa leprieurii and C.
RT   apomeiotica.";
RL   J. Phycol. 34:361-370(1998).
DR   EMBL; D89960; BAA31305.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             26.4%;  Score 14;  DB 10;  Length 9;
  Best Local Similarity   22.2%;  Pred. No. 3e+05;
  Matches    2;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 YLHDPEFNL 9
           :: |  |:
Db       1 FVETPTANV 9
```

```
RESULT  20
O81968
ID   O81968      PRELIMINARY;      PRT;     9 AA.
AC   O81968;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa ogasawaraensis.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=596;
RA   Kamiya M., West J.A., King R.J., Zuccarello G.C., Tanaka J., Hara Y.;
RT   "Evolutionary divergence in the red algae Caloglossa leprieurii and C.
RT   apomeiotica.";
RL   J. Phycol. 34:361-370(1998).
DR   EMBL; D89961; BAA31307.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             26.4%;  Score 14;  DB 10;  Length 9;
  Best Local Similarity   22.2%;  Pred. No. 3e+05;
  Matches    2;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 YLHDPEFNL 9
           :: |  |:
Db       1 FVETPTANV 9


RESULT  21
O82778
ID   O82778      PRELIMINARY;      PRT;     9 AA.
AC   O82778;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa leprieurii.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1048, 902, 490, 932, 922, 880, 1053, 1052;
RA   Kamiya M., West J.A., King R.J., Zuccarello G.C., Tanaka J., Hara Y.;
RT   "Evolutionary divergence in the red algae Caloglossa leprieurii and C.
RT   apomeiotica.";
RL   J. Phycol. 34:361-370(1998).
DR   EMBL; D89959; BAA31303.1; -.
DR   EMBL; D89949; BAA31283.1; -.
DR   EMBL; D89951; BAA31287.1; -.
DR   EMBL; D89952; BAA31289.1; -.
DR   EMBL; D89953; BAA31291.1; -.
DR   EMBL; D89954; BAA31293.1; -.
DR   EMBL; D89955; BAA31295.1; -.
DR   EMBL; D89956; BAA31297.1; -.
DR   EMBL; D89957; BAA31299.1; -.
DR   EMBL; D89958; BAA31301.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             26.4%;  Score 14;  DB 10;  Length 9;
  Best Local Similarity   22.2%;  Pred. No. 3e+05;
  Matches    2;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy       1 YLHDPEFNL 9
           ::  |  |:
Db       1 FVETPTANV 9


RESULT   22
O99182
ID   O99182      PRELIMINARY;      PRT;     7 AA.
AC   O99182;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CYTOCHROME OXIDASE I (FRAGMENT).
GN   COI.
OS   Pterolebias zonatus.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
OC   Atherinomorpha; Cyprinodontiformes; Aplocheiloidei; Aplocheilidae;
OC   Pterolebias.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Murphy W.J., Thomerson J.E., Collier G.E.;
RT   "A molecular phylogeny of Neotropical aplocheiloid killifishes
RT   (Cyprinodontiformes, Rivulidae).";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF002591; AAD01074.1; -.
KW   Mitochondrion.
FT   NON_TER       1      1
SQ   SEQUENCE   7 AA;  899 MW;  672721F6CB572030 CRC64;


  Query Match             24.5%;  Score 13;  DB 8;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLH 3
           | |
Db       3 YQH 5


RESULT   23
Q9R7T2
ID   Q9R7T2      PRELIMINARY;      PRT;     8 AA.
AC   Q9R7T2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   HYPOTHETICAL PROTEIN HI0004 (FRAGMENT).
GN   YQFG.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA Sequence of Escherichia coli K-12 Genome Corresponding
RT   to the 12.7-28.0 min Region on the Linkage Map.";
RL   DNA Res. 3:137-155(1996).
DR   EMBL; D90705; BAA35310.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  964 MW;  DF133B1DD04B476A CRC64;


  Query Match             24.5%;  Score 13;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 DP 5
           ||
Db       1 DP 2


RESULT   24
Q9UMH9
ID   Q9UMH9      PRELIMINARY;      PRT;     8 AA.
AC   Q9UMH9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RHCE PROTEIN (FRAGMENT).
GN   RHCE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RA   Matassi G., Cherif-Zahar B., Mouro I., Cartron J.P.;
RT   "Characterization of the recombination hot spot involved in the
RT   genomic rearrangement leading to the hybrid D-CE-D gene in the DVI
RT   phenotype.";
RL   Am. J. Hum. Genet. 60:808-817(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE; 90349591.
RA   Cherif-Zahar B., Bloy C., Le Van Kim C., Blanchard D., Bailly P.,
RA   Hermand P., Salmon C., Cartron J.-P., Colin Y.;
RT   "Molecular cloning and protein structure of a human blood group Rh
RT   polypeptide.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:6243-6247(1990).
DR   EMBL; Z97030; CAB09726.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  1049 MW;  C007244691FB5AB1 CRC64;


  Query Match             24.5%;  Score 13;  DB 4;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 DPEFNL 9
           |   ||
Db       2 DYHMNL 7


RESULT   25
Q9TT78
ID   Q9TT78      PRELIMINARY;      PRT;     8 AA.
AC   Q9TT78;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   THYMIDYLATE SYNTHASE (FRAGMENT).
GN   TS.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Brouillette J.A., Andrew J.R., Venta P.J.;
RT   "Estimate of Nucleotide Diversity in Dogs Using a Pool-and-Sequence
RT   Method to Identify Single Nucleotide Polymorphisms.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF202073; AAF20918.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  899 MW;  6731A1E059CAA867 CRC64;
```

```
  Query Match             24.5%;  Score 13;  DB 6;  Length 8;
  Best Local Similarity   33.3%;  Pred. No. 3e+05;
  Matches    1;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLH 3
            ::|
Db        4 FIH 6


RESULT  26
Q51765
ID   Q51765      PRELIMINARY;      PRT;     9 AA.
AC   Q51765;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   UNKNOWN ORF (FRAGMENT).
OS   Pseudomonas fluorescens.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC   Pseudomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90132584.
RA   Benen J.A., Van Berkel W.J., Van Dongen W.M., Muller F., Kok A.;
RT   "Molecular cloning and sequence determination of the lpd gene encoding
RT   lipoamide dehydrogenase from Pseudomonas fluorescens.";
RL   J. Gen. Microbiol. 135:1787-1797(1989).
DR   EMBL; M28356; AAA99233.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1011 MW;  269EA72727340DC7 CRC64;


  Query Match             24.5%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        4 DP 5
            ||
Db        1 DP 2


RESULT  27
Q9R7H9
ID   Q9R7H9      PRELIMINARY;      PRT;     9 AA.
AC   Q9R7H9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   LIPOPROTEIN (FRAGMENT).
GN   NLPD.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98083063.
RA   Martin K., Morlin G., Smith A., Nordyke A., Eisenstark A., Golomb M.;
RT   "The tryptophanase gene cluster of Haemophilus influenzae type b:
RT   evidence for horizontal gene transfer.";
RL   J. Bacteriol. 180:107-118(1998).
DR   EMBL; AF003252; AAB96582.1; -.
KW   Lipoprotein.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1152 MW;  35A017673B4412D7 CRC64;


  Query Match             24.5%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        4 DP 5
            ||
Db        1 DP 2


RESULT  28
Q9R735
ID   Q9R735      PRELIMINARY;      PRT;     9 AA.
AC   Q9R735;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   FKBA PROTEIN (FRAGMENT).
GN   FKBA.
OS   Streptomyces chrysomallus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94341259.
RA   Pahl A., Keller U.;
RT   "Streptomyces chrysomallus FKBP-33 is a novel immunophilin consisting
RT   of two FK506 binding domains; its gene is transcriptionally coupled to
RT   the FKBP-12 gene.";
RL   EMBO J. 13:3472-3480(1994).
DR   EMBL; Z34523; CAA84282.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1047 MW;  9A8BCB07633B1045 CRC64;


  Query Match             24.5%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        2 LHDPE 6
             :  ||
Db        3 IEKPE 7


RESULT  29
P78484
ID   P78484      PRELIMINARY;      PRT;     9 AA.
AC   P78484;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   FYNC ONCOGENE, 3' END, CLONE PFYN-C-11. (FRAGMENT).
GN   FYN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE OF 90-98 FROM N.A.
RX   MEDLINE; 88234523.
RA   Kawakami T., Kawakami Y., Aaronson S.A., Robbins K.C.;
RT   "Acquisition of transforming properties by FYN, a normal SRC-related
RT   human gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:3870-3874(1988).
DR   EMBL; M20284; AAA52491.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1143 MW;  12BA1729D5A6D73B CRC64;


  Query Match             24.5%;  Score 13;  DB 4;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLHD 4
            :| |
Db        6 FLED 9


RESULT  30
Q9UM87
ID   Q9UM87      PRELIMINARY;      PRT;     9 AA.
```

```
AC   Q9UM87;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   FGFR2 PROTEIN (FRAGMENT).
GN   FGFR2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96253074.
RA   Wada C., Ishigaki M., Toyo-oka Y., Yamabe H., Ohnuki Y., Takada F.,
RA   Yamazaki Y., Ohtani H.;
RT   "[Nucleotide sequences at intron 6 and exon 7 junction of fibroblast
RT   growth factor receptor 2 and rapid mutational analysis in Apert
RT   syndrome].";
RL   Rinsho Byori 44:435-438(1996).
DR   EMBL; S82438; AAD14392.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1103 MW;  9E4D20477401F775 CRC64;


  Query Match             24.5%;  Score 13;  DB 4;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 HDP 5
           | |
Db       4 HRP 6


RESULT  31
Q9UC36
ID   Q9UC36      PRELIMINARY;      PRT;     9 AA.
AC   Q9UC36;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   28 KDA HEAT SHOCK PROTEIN HOMOLOG FRAGMENT 1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92218434.
RA   Kato K., Shinohara H., Goto S., Inaguma Y., Morishita R., Asano T.;
RT   "Copurification of small heat shock protein with alpha B crystallin
RT   from human skeletal muscle.";
RL   J. Biol. Chem. 267:7718-7725(1992).
SQ   SEQUENCE   9 AA;  1220 MW;  26933415B1F77B43 CRC64;


  Query Match             24.5%;  Score 13;  DB 4;  Length 9;
  Best Local Similarity   42.9%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 YLHDPEF 7
           | |   |
Db       3 YPHSRLF 9


RESULT  32
Q9TWD6
ID   Q9TWD6      PRELIMINARY;      PRT;     9 AA.
AC   Q9TWD6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   LED-NPF-1=NEUROPEPTIDE F-RELATED PEPTIDE.
OS   Leptinotarsa decemlineata (Colorado potato beetle).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC   Cucujiformia; Chrysomelidae; Chrysomelinae; Leptinotarsa.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 96245438.
RA   Spittaels K., Verhaert P., Shaw C., Johnston R.N., Devreese B.,
RA   Van Beeumen J., De Loof A.;
RT   "Insect neuropeptide F (NPF)-related peptides: isolation from Colorado
RT   potato beetle (Leptinotarsa decemlineata) brain.";
RL   Insect Biochem. Mol. Biol. 26:375-382(1996).
SQ   SEQUENCE   9 AA;  1066 MW;  7E02340736C76864 CRC64;


  Query Match             24.5%;  Score 13;  DB 5;  Length 9;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 PEFNL 9
           |:  |
Db       4 PQLRL 8


RESULT  33
Q9XJN0
ID   Q9XJN0      PRELIMINARY;      PRT;     9 AA.
AC   Q9XJN0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   P10 (FRAGMENT).
OS   bacteriophage phi-10.
OC   Viruses; dsRNA viruses; Cystoviridae; Cystovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mindich L., Qiao X., Qiao J., Romantschuk M., Hoogstraten D.;
RT   "Isolation of additional bacteriophages with genomes of segmented
RT   double-stranded RNA.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF125675; AAD22555.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1058 MW;  880E376AA720544A CRC64;


  Query Match             24.5%;  Score 13;  DB 9;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 DP 5
           ||
Db       6 DP 7


RESULT  34
Q85723
ID   Q85723      PRELIMINARY;      PRT;     9 AA.
AC   Q85723;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   SIS ONCOGENE (FRAGMENT).
OS   Simian sarcoma virus.
OC   Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84106822.
RA   Devare S.G., Shatzman A.R., Robbins K.C., Rosenberg M., Aaronson S.A.;
RT   "Expression of the PDGF-related transforming protein of simian sarcoma
RT   virus in E. coli.";
RL   Cell 36:43-49(1984).
DR   EMBL; K01473; AAA46816.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1048 MW;  9C53A866C361A731 CRC64;
```

```
  Query Match             24.5%;  Score 13;  DB 12;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        4 DP 5
            ||
Db        8 DP 9


RESULT  35
P82099
ID   P82099     PRELIMINARY;      PRT;     5 AA.
AC   P82099;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ELECTRIN 3.
OS   Litoria rubella (Desert tree frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Hylidae;
OC   Litoria.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=SKIN SECRETION;
RA   Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;
RT   "Peptides from the skin glands of the Australian buzzing tree frog
RT   Litori electrica. Comparison with the skin peptides from Litoria
RT   rubella.";
RL   Aust. J. Chem. 52:0-0(1999).
KW   Amphibian skin; Amidation.
FT   MOD_RES       5      5       AMIDATION.
SQ   SEQUENCE   5 AA;  630 MW;  668761F2C9A00000 CRC64;


  Query Match             22.6%;  Score 12;  DB 13;  Length 5;
  Best Local Similarity   33.3%;  Pred. No. 3e+05;
  Matches    1;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLH 3
            ::|
Db        1 FVH 3


Search completed: December 16, 2000, 04:22:14
Job time:`4608 sec
```

```
                         GenCore version 4.5
                 Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 00:51:16 ; Search time 108.84 Seconds
                                               (without alignments)
                                               2.827 Million cell updates/sec

Title:           US-09-529-121-10
Perfect score:   53
Sequence:        1 YLHDPEFNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters:      268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_36:*
                1:  /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
                2:  /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
                3:  /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
                4:  /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
                5:  /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
                6:  /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
                7:  /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
                8:  /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
                9:  /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
               10:  /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
               11:  /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
               12:  /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
               13:  /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
               14:  /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
               15:  /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
               16:  /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
               17:  /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
               18:  /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
               19:  /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
               20:  /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
               21:  /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
               %
Result         Query
  No.  Score   Match Length DB  ID                      Description
 ----------------------------------------------------------------------
    1     53   100.0   2183 14  R39592                  L protein of atten
    2     53   100.0   2183 19  W48706                  Measles virus Mont
    3     53   100.0   2183 19  W48707                  Measles virus Rube
    4     53   100.0   2183 19  W48708                  Measles virus Mora
    5     53   100.0   2183 19  W48709                  Measles virus Zagr
    6     53   100.0   2183 19  W48710                  Measles virus AIK-
    7     53   100.0   2183 19  W48703                  Measles virus Edmo
    8     53   100.0   2183 19  W48704                  Measles virus 1977
    9     53   100.0   2183 19  W48705                  Measles virus 1983
   10     37    69.8    329 15  R51284                  L-fucose dehydroge
   11     37    69.8   2233 19  W48711                  HPIV-3 JS isolate
   12     37    69.8   2233 19  W48712                  HPIV-3 FRhl cp45 v
   13     37    69.8   2233 19  W48713                  HPIV-3 Vero cp45 v
   14     36    67.9   2343 20  W80989                  Canine factor VIII
   15     36    67.9   2343 21  Y57846                  Canine Factor VIII
   16     35    66.0    693 17  R89114                  Bombyx mori (pro)p
   17     35    66.0    950 20  W83319                  Mouse SMAD interac
   18     34    64.2    386 12  R10624                  DNA methylase enco
   19     34    64.2   3139 20  Y08898                  Human Huntington's
   20   33.5    63.2    919 18  W15197                  Human EHOC-17 prot
   21     33    62.3    659 18  W24123                  Protease.  Synthet
   22     33    62.3    659 18  W24121                  Thermococcus prote
   23     33    62.3    659 20  W94840                  WO9856926 Seq ID 1
   24     33    62.3   1151 21  Y83251                  F-box protein Grrl
   25     33    62.3   1230 19  W48895                  Candida albicans C
   26     32    60.4    321 13  R27118                  Fucose dehydrogena
   27     32    60.4    339 20  Y37246                  Amino acid sequenc
   28     32    60.4    750 20  W88407                  Human adult neural
   29     32    60.4    906 20  Y37168                  Protein involved i
   30     31    58.5     22 20  W96828                  Nucleic acid bindi
   31     31    58.5     33 20  Y06875                  Minimised protein
   32     31    58.5     33 21  Y77602                  Staphylococcal pro
   33     31    58.5     33 21  Y77616                  Staphylococcal pro
   34     31    58.5     33 21  Y77626                  Staphylococcal pro
   35     31    58.5     34 20  Y06879                  Sequence of synthe
   36     31    58.5     34 21  Y77604                  Staphylococcal pro
   37     31    58.5     34 21  Y77637                  Staphylococcal pro
   38     31    58.5     34 21  Y77647                  Staphylococcal pro
   39     31    58.5     38 21  Y77606                  Staphylococcal pro
   40     31    58.5     38 21  Y77607                  Staphylococcal pro
   41     31    58.5     38 21  Y77608                  Staphylococcal pro
   42     31    58.5     38 21  Y77609                  Staphylococcal pro
   43     31    58.5     38 21  Y77610                  Staphylococcal pro
   44     31    58.5     38 21  Y77611                  Staphylococcal pro
   45     31    58.5     38 21  Y77612                  Staphylococcal pro

                                   ALIGNMENTS

RESULT   1
R39592
ID   R39592 standard; Protein; 2183 AA.
XX
AC   R39592;
XX
DT   19-AUG-1993  (first entry)
XX
DE   L protein of attenuated measles virus strain AIK-C.
XX
KW   paramyxoviridae; RNA virus; attenuation; vaccine.
XX
OS   Human measles virus.
XX
PN   EP540135-A.
XX
PD   05-MAY-1993.
XX
PF   10-MAR-1992;   92EP-0302004.
XX
PR   14-OCT-1991;   91JP-0293625.
XX
PA   (KITA ) KITASATO INST.
PA   (KITA ) KITASATO KENKYUSHO SH.
XX
PI   Makino S,  Mori T,  Sasaki K;
XX
DR   WPI; 1993-145503/18.
DR   N-PSDB; Q40480.
XX
PT   New attenuated measles vaccine virus strain - retains high
PT   immunogenicity with reduced pyrogenicity and having no
PT   neurological complications
XX
PS   Disclosure; Page 22-37; 47pp; English.
```

```
XX
CC   The sequence of the entire genome of the attenuated measles
CC   virus strain AIK-C has been determined and contains 6 open reading
CC   frames. The sequence of the L protein was deduced from the sixth
CC   ORF. The virus retains high immunogenicity with reduced pyrogenicity
CC   and no neurological complications.
CC   See also R34537-R34541.
XX
SQ   Sequence   2183 AA;


  Query Match             100.0%;  Score 53;  DB 14;  Length 2183;
  Best Local Similarity   100.0%;  Pred. No. 0.16;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLHDPEFNL 9
             |||||||||
Db     522 ylhdpefnl 530


RESULT   2
W48706
ID   W48706 standard; Protein; 2183 AA.
XX
AC   W48706;
XX
DT   13-OCT-1998  (first entry)
XX
DE   Measles virus Montefiore isolate wild-type L protein.
XX
KW   L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW   single stranded RNA virus; Mononegavirales.
XX
OS   Measles virus.
XX
PN   WO9813501-A2.
XX
PD   02-APR-1998.
XX
PF   19-SEP-1997;   97WO-US16718.
XX
PR   27-SEP-1996;   96US-0026823.
XX
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Murphy BR,  Randolph VB,  Sidhu MS,  Tatem JM,  Udem SA;
XX
DR   WPI; 1998-230710/20.
DR   N-PSDB; V18267.
XX
PT   Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT   single stranded RNA virus of order Mononegavirales  - having
PT   attenuating mutation in 3' genomic promoter region and RNA
PT   polymerase gene, useful as vaccine to immunise against such virus
XX
PS   Disclosure; Page 153-161; 426pp; English.
XX
CC   This sequence represents the wild-type L protein from Measles virus
CC   isolate Montefiore. This sequence is used in a method which involves the
CC   isolation of recombinantly-generated, attenuated, non-segmented,
CC   negative-sense, single stranded RNA virus of the order Mononegavirales
CC   which have at least 1 attenuating mutation in the 3' genomic promoter
CC   region and at least 1 attenuating mutation in the RNA polymerase gene.
CC   This RNA virus can be used as a vaccine to immunise an individual against
CC   such a virus.
XX
SQ   Sequence   2183 AA;


  Query Match             100.0%;  Score 53;  DB 19;  Length 2183;
  Best Local Similarity   100.0%;  Pred. No. 0.16;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLHDPEFNL 9
             |||||||||
Db     522 ylhdpefnl 530


RESULT   3
W48707
ID   W48707 standard; Protein; 2183 AA.
XX
AC   W48707;
XX
DT   13-OCT-1998  (first entry)
XX
DE   Measles virus Rubeovax vaccine L protein.
XX
KW   L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW   single stranded RNA virus; Mononegavirales.
XX
OS   Measles virus.
XX
PN   WO9813501-A2.
XX
PD   02-APR-1998.
XX
PF   19-SEP-1997;   97WO-US16718.
XX
PR   27-SEP-1996;   96US-0026823.
XX
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Murphy BR,  Randolph VB,  Sidhu MS,  Tatem JM,  Udem SA;
XX
DR   WPI; 1998-230710/20.
DR   N-PSDB; V18268.
XX
PT   Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT   single stranded RNA virus of order Mononegavirales  - having
PT   attenuating mutation in 3' genomic promoter region and RNA
PT   polymerase gene, useful as vaccine to immunise against such virus
XX
PS   Disclosure; Page 172-180; 426pp; English.
XX
CC   This sequence represents the L protein from Measles virus Rubeovax
CC   vaccine. This sequence is used in a method which involves the
CC   isolation of recombinantly-generated, attenuated, non-segmented,
CC   negative-sense, single stranded RNA virus of the order Mononegavirales
CC   which have at least 1 attenuating mutation in the 3' genomic promoter
CC   region and at least 1 attenuating mutation in the RNA polymerase gene.
CC   This RNA virus can be used as a vaccine to immunise an individual against
CC   such a virus.
XX
SQ   Sequence   2183 AA;


  Query Match             100.0%;  Score 53;  DB 19;  Length 2183;
  Best Local Similarity   100.0%;  Pred. No. 0.16;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLHDPEFNL 9
             |||||||||
Db     522 ylhdpefnl 530


RESULT   4
W48708
ID   W48708 standard; Protein; 2183 AA.
XX
AC   W48708;
XX
DT   13-OCT-1998  (first entry)
XX
```

```
DE   Measles virus Moraten vaccine L protein.
XX
KW   L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW   single stranded RNA virus; Mononegavirales.
XX
OS   Measles virus.
XX
PN   WO9813501-A2.
XX
PD   02-APR-1998.
XX
PF   19-SEP-1997;   97WO-US16718.
XX
PR   27-SEP-1996;   96US-0026823.
XX
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Murphy BR,  Randolph VB,  Sidhu MS,  Tatem JM,  Udem SA;
XX
DR   WPI; 1998-230710/20.
DR   N-PSDB; W48708.
XX
PT   Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT   single stranded RNA virus of order Mononegavirales  - having
PT   attenuating mutation in 3' genomic promoter region and RNA
PT   polymerase gene, useful as vaccine to immunise against such virus
XX
PS   Disclosure; Page 190-198; 426pp; English.
XX
CC   This sequence represents the L protein from Measles virus Moraten
CC   vaccine. This sequence is used in a method which involves the
CC   isolation of recombinantly-generated, attenuated, non-segmented,
CC   negative-sense, single stranded RNA virus of the order Mononegavirales
CC   which have at least 1 attenuating mutation in the 3' genomic promoter
CC   region and at least 1 attenuating mutation in the RNA polymerase gene.
CC   This RNA virus can be used as a vaccine to immunise an individual against
CC   such a virus.
XX
SQ   Sequence   2183 AA;

  Query Match             100.0%;  Score 53;  DB 19;  Length 2183;
  Best Local Similarity   100.0%;  Pred. No. 0.16;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YLHDPEFNL 9
          |||||||||
Db    522 ylhdpefnl 530


RESULT   5
W48709
ID   W48709 standard; Protein; 2183 AA.
XX
AC   W48709;
XX
DT   13-OCT-1998  (first entry)
XX
DE   Measles virus Zagreb vaccine L protein.
XX
KW   L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW   single stranded RNA virus; Mononegavirales.
XX
OS   Measles virus.
XX
PN   WO9813501-A2.
XX
PD   02-APR-1998.
XX
PF   19-SEP-1997;   97WO-US16718.
XX
PR   27-SEP-1996;   96US-0026823.
```

```
XX
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Murphy BR,  Randolph VB,  Sidhu MS,  Tatem JM,  Udem SA;
XX
DR   WPI; 1998-230710/20.
DR   N-PSDB; V18270.
XX
PT   Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT   single stranded RNA virus of order Mononegavirales  - having
PT   attenuating mutation in 3' genomic promoter region and RNA
PT   polymerase gene, useful as vaccine to immunise against such virus
XX
PS   Disclosure; Page 209-217; 426pp; English.
XX
CC   This sequence represents the L protein from Measles virus Zagreb vaccine.
CC   This sequence is used in a method which involves the isolation of
CC   recombinantly-generated, attenuated, non-segmented, negative-sense,
CC   single stranded RNA virus of the order Mononegavirales which have at
CC   least 1 attenuating mutation in the 3' genomic promoter region and at
CC   least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
CC   can be used as a vaccine to immunise an individual against such a virus.
XX
SQ   Sequence   2183 AA;

  Query Match             100.0%;  Score 53;  DB 19;  Length 2183;
  Best Local Similarity   100.0%;  Pred. No. 0.16;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YLHDPEFNL 9
          |||||||||
Db    522 ylhdpefnl 530


RESULT   6
W48710
ID   W48710 standard; Protein; 2183 AA.
XX
AC   W48710;
XX
DT   13-OCT-1998  (first entry)
XX
DE   Measles virus AIK-C vaccine L protein.
XX
KW   L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW   single stranded RNA virus; Mononegavirales.
XX
OS   Measles virus.
XX
PN   WO9813501-A2.
XX
PD   02-APR-1998.
XX
PF   19-SEP-1997;   97WO-US16718.
XX
PR   27-SEP-1996;   96US-0026823.
XX
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Murphy BR,  Randolph VB,  Sidhu MS,  Tatem JM,  Udem SA;
XX
DR   WPI; 1998-230710/20.
DR   N-PSDB; V18271.
XX
PT   Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT   single stranded RNA virus of order Mononegavirales  - having
PT   attenuating mutation in 3' genomic promoter region and RNA
PT   polymerase gene, useful as vaccine to immunise against such virus
XX
PS   Disclosure; Page 228-236; 426pp; English.
```

```
XX
CC   This sequence represents the L protein from Measles virus AIK-C vaccine.
CC   This sequence is used in a method which involves the isolation of
CC   recombinantly-generated, attenuated, non-segmented, negative-sense,
CC   single stranded RNA virus of the order Mononegavirales which have at
CC   least 1 attenuating mutation in the 3' genomic promoter region and at
CC   least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
CC   can be used as a vaccine to immunise an individual against such a virus.
XX
SQ   Sequence   2183 AA;

  Query Match             100.0%;  Score 53;  DB 19;  Length 2183;
  Best Local Similarity   100.0%;  Pred. No. 0.16;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YLHDPEFNL 9
          |||||||||
Db    522 ylhdpefnl 530


RESULT   7
W48703
ID   W48703 standard; Protein; 2183 AA.
XX
AC   W48703;
XX
DT   13-OCT-1998  (first entry)
XX
DE   Measles virus Edmonston isolate wild-type L protein.
XX
KW   L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW   single stranded RNA virus; Mononegavirales.
XX
OS   Measles virus.
XX
PN   WO9813501-A2.
XX
PD   02-APR-1998.
XX
PF   19-SEP-1997;   97WO-US16718.
XX
PR   27-SEP-1996;   96US-0026823.
XX
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Murphy BR,  Randolph VB,  Sidhu MS,  Tatem JM,  Udem SA;
XX
DR   WPI; 1998-230710/20.
DR   N-PSDB; V18264.
XX
PT   Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT   single stranded RNA virus of order Mononegavirales  - having
PT   attenuating mutation in 3' genomic promoter region and RNA
PT   polymerase gene, useful as vaccine to immunise against such virus
XX
PS   Disclosure; Page 97-105; 426pp; English.
XX
CC   This sequence represents a wild-type L protein from Measles virus isolate
CC   Edmonston. This sequence is used in a method which involves the isolation
CC   of recombinantly-generated, attenuated, non-segmented, negative-sense,
CC   single stranded RNA virus of the order Mononegavirales which have at
CC   least 1 attenuating mutation in the 3' genomic promoter region and at
CC   least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
CC   can be used as a vaccine to immunise an individual against such a virus.
XX
SQ   Sequence   2183 AA;

  Query Match             100.0%;  Score 53;  DB 19;  Length 2183;
  Best Local Similarity   100.0%;  Pred. No. 0.16;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YLHDPEFNL 9
          |||||||||
Db    522 ylhdpefnl 530


RESULT   8
W48704
ID   W48704 standard; Protein; 2183 AA.
XX
AC   W48704;
XX
DT   13-OCT-1998  (first entry)
XX
DE   Measles virus 1977 isolate L protein.
XX
KW   L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW   single stranded RNA virus; Mononegavirales.
XX
OS   Measles virus.
XX
PN   WO9813501-A2.
XX
PD   02-APR-1998.
XX
PF   19-SEP-1997;   97WO-US16718.
XX
PR   27-SEP-1996;   96US-0026823.
XX
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Murphy BR,  Randolph VB,  Sidhu MS,  Tatem JM,  Udem SA;
XX
DR   WPI; 1998-230710/20.
DR   N-PSDB; V18265.
XX
PT   Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT   single stranded RNA virus of order Mononegavirales  - having
PT   attenuating mutation in 3' genomic promoter region and RNA
PT   polymerase gene, useful as vaccine to immunise against such virus
XX
PS   Disclosure; Page 116-124; 426pp; English.
XX
CC   This sequence represents the L protein from Measles virus isolate 1977.
CC   This sequence is used in a method which involves the isolation
CC   of recombinantly-generated, attenuated, non-segmented, negative-sense,
CC   single stranded RNA virus of the order Mononegavirales which have at
CC   least 1 attenuating mutation in the 3' genomic promoter region and at
CC   least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
CC   can be used as a vaccine to immunise an individual against such a virus.
XX
SQ   Sequence   2183 AA;

  Query Match             100.0%;  Score 53;  DB 19;  Length 2183;
  Best Local Similarity   100.0%;  Pred. No. 0.16;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YLHDPEFNL 9
          |||||||||
Db    522 ylhdpefnl 530


RESULT   9
W48705
ID   W48705 standard; Protein; 2183 AA.
XX
AC   W48705;
XX
DT   13-OCT-1998  (first entry)
XX
DE   Measles virus 1983 isolate wild-type L protein.
```

```
XX
KW   L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW   single stranded RNA virus; Mononegavirales.
XX
OS   Measles virus.
XX
PN   WO9813501-A2.
XX
PD   02-APR-1998.
XX
PF   19-SEP-1997;   97WO-US16718.
XX
PR   27-SEP-1996;   96US-0026823.
XX
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Murphy BR,  Randolph VB,  Sidhu MS,  Tatem JM,  Udem SA;
XX
DR   WPI; 1998-230710/20.
DR   N-PSDB; V18266.
XX
PT   Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT   single stranded RNA virus of order Mononegavirales  - having
PT   attenuating mutation in 3' genomic promoter region and RNA
PT   polymerase gene, useful as vaccine to immunise against such virus
XX
PS   Disclosure; Page 134-142; 426pp; English.
XX
CC   This sequence represents a wild-type L protein from Measles virus isolate
CC   1983. This sequence is used in a method which involves the isolation
CC   of recombinantly-generated, attenuated, non-segmented, negative-sense,
CC   single stranded RNA virus of the order Mononegavirales which have at
CC   least 1 attenuating mutation in the 3' genomic promoter region and at
CC   least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
CC   can be used as a vaccine to immunise an individual against such a virus.
XX
SQ   Sequence   2183 AA;


  Query Match             100.0%;  Score 53;  DB 19;  Length 2183;
  Best Local Similarity   100.0%;  Pred. No. 0.16;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YLHDPEFNL 9
          |||||||||
Db    522 ylhdpefnl 530


RESULT  10
R51284
ID   R51284 standard; Protein; 329 AA.
XX
AC   R51284;
XX
DT   14-DEC-1994  (first entry)
XX
DE   L-fucose dehydrogenase.
XX
KW   L-fucose dehydrogenase; L-FDH; enzyme.
XX
OS   Pseudomonas sp. No.1143.
XX
PN   JP06090765-A.
XX
PD   05-APR-1994.
XX
PF   11-SEP-1992;   92JP-0243372.
XX
PR   11-SEP-1992;   92JP-0243372.
XX
PA   (KIKK ) KIKKOMAN CORP.
PA   (NODA ) NODA INST SCI RES.
```

```
PA   (NODA ) ZH NODA SANGYO KAGAKU KENKYUSHO.
XX
PI   Horiuchi T,  Koyama Y,  Nakano E,  Otake H;
XX
DR   WPI; 1994-146990/18.
DR   N-PSDB; Q62089.
XX
PT   Novel L-fucose dehydrogenase gene - a new recombinant DNA and the
PT   prepn. of L-FDH using a L-FDH vector
XX
PS   Claim 1; Page 5; 6pp; Japanese.
XX
CC   The L-FDH gene is new and can be used for the recombinant prodn. of
CC   the enzyme.
XX
SQ   Sequence   329 AA;

  Query Match             69.8%;  Score 37;  DB 15;  Length 329;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YLHDPE 6
          ||||||
Db    143 ylhdpe 148


RESULT  11
W48711
ID   W48711 standard; Protein; 2233 AA.
XX
AC   W48711;
XX
DT   13-OCT-1998  (first entry)
XX
DE   HPIV-3 JS isolate wild-type L protein.
XX
KW   L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW   single stranded RNA virus; Mononegavirales.
XX
OS   Human parainfluenza virus.
XX
PN   WO9813501-A2.
XX
PD   02-APR-1998.
XX
PF   19-SEP-1997;   97WO-US16718.
XX
PR   27-SEP-1996;   96US-0026823.
XX
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Murphy BR,  Randolph VB,  Sidhu MS,  Tatem JM,  Udem SA;
XX
DR   WPI; 1998-230710/20.
DR   N-PSDB; V18272.
XX
PT   Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT   single stranded RNA virus of order Mononegavirales  - having
PT   attenuating mutation in 3' genomic promoter region and RNA
PT   polymerase gene, useful as vaccine to immunise against such virus
XX
PS   Disclosure; Page 246-254; 426pp; English.
XX
CC   This sequence represents the wild-type L protein from Human parainfluenza
CC   virus (HPIV-3) type 3 isolate JS. This sequence is used in a method which
CC   involves the isolation of recombinantly-generated, attenuated,
CC   non-segmented, negative-sense, single stranded RNA virus of the order
CC   Mononegavirales which have at least 1 attenuating mutation in the 3'
CC   genomic promoter region and at least 1 attenuating mutation in the RNA
CC   polymerase gene. This RNA virus can be used as a vaccine to immunise an
CC   individual against such a virus.
```

```
XX
SQ   Sequence   2233 AA;


  Query Match             69.8%;  Score 37;  DB 19;  Length 2233;
  Best Local Similarity   66.7%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLHDPEFNL 9
          :| |||||:
Db    530 wlddpefni 538


RESULT  12
W48712
ID   W48712 standard; Protein; 2233 AA.
XX
AC   W48712;
XX
DT   13-OCT-1998  (first entry)
XX
DE   HPIV-3 FRhl cp45 vaccine L protein.
XX
KW   L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW   single stranded RNA virus; Mononegavirales.
XX
OS   Human parainfluenza virus.
XX
PN   WO9813501-A2.
XX
PD   02-APR-1998.
XX
PF   19-SEP-1997;   97WO-US16718.
XX
PR   27-SEP-1996;   96US-0026823.
XX
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Murphy BR,  Randolph VB,  Sidhu MS,  Tatem JM,  Udem SA;
XX
DR   WPI; 1998-230710/20.
DR   N-PSDB; V18273.
XX
PT   Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT   single stranded RNA virus of order Mononegavirales  - having
PT   attenuating mutation in 3' genomic promoter region and RNA
PT   polymerase gene, useful as vaccine to immunise against such virus
XX
PS   Disclosure; Page 265-273; 426pp; English.
XX
CC   This sequence represents the Human parainfluenza virus (HPIV-3) type 3
CC   vaccine FRhl cp45 L protein. This sequence is used in a method which
CC   involves the isolation of recombinantly-generated, attenuated,
CC   non-segmented, negative-sense, single stranded RNA virus of the order
CC   Mononegavirales which have at least 1 attenuating mutation in the 3'
CC   genomic promoter region and at least 1 attenuating mutation in the RNA
CC   polymerase gene. This RNA virus can be used as a vaccine to immunise an
CC   individual against such a virus.
XX
SQ   Sequence   2233 AA;


  Query Match             69.8%;  Score 37;  DB 19;  Length 2233;
  Best Local Similarity   66.7%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLHDPEFNL 9
          :| |||||:
Db    530 wlddpefni 538


RESULT  13
W48713
ID   W48713 standard; Protein; 2233 AA.
XX
AC   W48713;
XX
DT   13-OCT-1998  (first entry)
XX
DE   HPIV-3 Vero cp45 vaccine L protein.
XX
KW   L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW   single stranded RNA virus; Mononegavirales.
XX
OS   Human parainfluenza virus.
XX
PN   WO9813501-A2.
XX
PD   02-APR-1998.
XX
PF   19-SEP-1997;   97WO-US16718.
XX
PR   27-SEP-1996;   96US-0026823.
XX
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Murphy BR,  Randolph VB,  Sidhu MS,  Tatem JM,  Udem SA;
XX
DR   WPI; 1998-230710/20.
DR   N-PSDB; V18274.
XX
PT   Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT   single stranded RNA virus of order Mononegavirales  - having
PT   attenuating mutation in 3' genomic promoter region and RNA
PT   polymerase gene, useful as vaccine to immunise against such virus
XX
PS   Disclosure; Page 283-291; 426pp; English.
XX
CC   This sequence represents the Human parainfluenza virus (HPIV-3) type 3
CC   vaccine Vero cp45 L protein. This sequence is used in a method which
CC   involves the isolation of recombinantly-generated, attenuated,
CC   non-segmented, negative-sense, single stranded RNA virus of the order
CC   Mononegavirales which have at least 1 attenuating mutation in the 3'
CC   genomic promoter region and at least 1 attenuating mutation in the RNA
CC   polymerase gene. This RNA virus can be used as a vaccine to immunise an
CC   individual against such a virus.
XX
SQ   Sequence   2233 AA;


  Query Match             69.8%;  Score 37;  DB 19;  Length 2233;
  Best Local Similarity   66.7%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLHDPEFNL 9
          :| |||||:
Db    530 wlddpefni 538


RESULT  14
W80989
ID   W80989 standard; Protein; 2343 AA.
XX
AC   W80989;
XX
DT   10-MAY-1999  (first entry)
XX
DE   Canine factor VIII.
XX
KW   Factor VIII; canine; dog; diagnosis; animal model; haemophilia A;
KW   gene therapy.
XX
OS   Canis familiaris.
XX
```

```
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /note= "signal peptide"
FT   Protein         20..2343
FT                   /note= "mature protein"
FT   Region          1664..1777
FT                   /note= "epitope-bearing region, specifically
FT                          claimed in Claim 24"
FT   Region          2278..2351
FT                   /note= "epitope-bearing region, specifically
FT                          claimed in Claim 24"
FT   Cleavage-site   385..386
FT                   /note= "thrombin cleavage site"
FT   Cleavage-site   753..754
FT                   /note= "thrombin cleavage site"
FT   Cleavage-site   1700..1701
FT                   /note= "thrombin cleavage site"
FT   Cleavage-site   350..351
FT                   /note= "protein C cleavage site"
FT   Cleavage-site   575..576
FT                   /note= "protein C cleavage site"
FT   Domain          1685..1695
FT                   /note= "A3 domain"
FT   Modified-site   359
FT                   /note= "sulfation"
FT   Modified-site   731
FT                   /note= "sulfation"
FT   Modified-site   732
FT                   /note= "sulfation"
FT   Modified-site   736
FT                   /note= "sulfation"
FT   Modified-site   1675
FT                   /note= "sulfation"
FT   Modified-site   1691
FT                   /note= "sulfation"
XX
PN   CA2225189-A.
XX
PD   06-SEP-1998.
XX
PF   06-MAR-1998;   98CA-2225189.
XX
PR   05-MAR-1998;   98US-0035141.
PR   06-MAR-1997;   97US-0039953.
XX
PA   (TOOH ) UNIV QUEENS KINGSTON.
XX
PI   Cameron C,  Horrocks L,  Hough C,  Lillicrap D,  Notley C;
XX
DR   WPI; 1999-071205/07.
DR   N-PSDB; V99801.
XX
PT   New canine factor VIII polynucleotide and polypeptide - useful for
PT   detection and treatment of haemophilia A using gene therapy
XX
PS   Claim 22; Fig 1; 153pp; English.
XX
CC   This polypeptide comprises canine factor VIII.  The canine Factor
CC   VIII gene (see V99801) was obtained by concatamerisation of
CC   RT-PCR-amplified factor VIII fragments obtained from canine liver
CC   total RNA.  A cDNA clone, designated pBK-cmV(1-6#23) canine FVIII,
CC   is deposited as ATCC 209475.  The invention provides canine factor
CC   VIII polypeptides encoded by such isolated nucleic acid molecules,
CC   antibodies binding to such polypeptides, genetic constructs
CC   comprising the nucleic acid molecules, prokaryotic or eukaryotic
CC   host cells, and methods and compositions for use in diagnosing and
CC   treating canine disorders characterised by factor VIII deficiency,
CC   especially haemophilia A.  A polynucleotide can be administered via
CC   a vector directly into canine cells (gene therapy).  The canine
CC   represents a potential model for gene therapy treatment of
CC   haemophilia in humans.  Epitope-bearing portions of canine factor
CC   VIII polypeptides can be used to generate anti-canine factor VII
CC   antibodies.
```

```
XX
SQ   Sequence   2343 AA;

  Query Match             67.9%;  Score 36;  DB 20;  Length 2343;
  Best Local Similarity   85.7%;  Pred. No. 2.7e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        3 HDPEFNL 9
            ||||| |
Db      617 hdpefql 623


RESULT  15
Y57846
ID   Y57846 standard; Protein; 2343 AA.
XX
AC   Y57846;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Canine Factor VIII (DOGLESS1) protein SEQ ID NO:2.
XX
KW   Canine; factor VIII; haemostatic; diagnosis; haemophilia A; dog.
XX
OS   Canis sp.
XX
PN   CA2264431-A1.
XX
PD   05-SEP-1999.
XX
PF   05-MAR-1999;   99CA-2264431.
XX
PR   05-MAR-1998;   98US-0035141.
PR   06-MAR-1998;   98CA-2225189.
XX
PA   (TOOH ) UNIV QUEENS KINGSTON.
XX
PI   Horrocks LSH,  Hough C,  Notley C,  Lillicrap D,  Cameron C;
XX
DR   WPI; 2000-073270/07.
DR   N-PSDB; Z56579.
XX
PT   Isolated nucleic acid encoding a canine factor VIII polypeptide for
PT   treating a disorder characterized by canine factor VIII deficiency,
PT   such as hemophilia A -
XX
PS   Claim 4; Fig 1; 152pp; English.
XX
CC   The present sequence represents canine factor VIII. The isolated
CC   factor VIII nucleic acid molecule and protein can be used for
CC   treating a disorder characterised by canine factor VIII
CC   deficiency in a canine, especially haemophilia A.
XX
SQ   Sequence   2343 AA;

  Query Match             67.9%;  Score 36;  DB 21;  Length 2343;
  Best Local Similarity   85.7%;  Pred. No. 2.7e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        3 HDPEFNL 9
            ||||| |
Db      617 hdpefql 623


Search completed: December 16, 2000, 00:51:17
Job time: 18777 sec
```

```
                        GenCore version 4.5
                 Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 04:23:32 ; Search time 68.03 Seconds
                                               (without alignments)
                                               4.227 Million cell updates/sec

Title:           US-09-529-121-10
Perfect score:   53
Sequence:        1 YLHDPEFNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters:      213

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       SwissProt_39:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                    %
Result             Query
  No.    Score     Match Length DB  ID                     Description
-------------------------------------------------------------------------------
    1      24      45.3      8  1  LCK1_LEUMA             P21140 leucophaea
    2      19      35.8      8  1  FAR1_PANRE             P41872 panagrellus
    3      19      35.8      8  1  LCK7_LEUMA             P19989 leucophaea
    4      19      35.8      9  1  FAR2_PANRE             P41873 panagrellus
    5      18      34.0      8  1  LCK2_LEUMA             P21141 leucophaea
    6      18      34.0      9  1  COXE_THUOB             P80975 thunnus obe
    7      17      32.1      8  1  ANG2_BOTJA             Q10582 bothrops ja
    8      16      30.2      8  1  ALL1_CYDPO             P82152 cydia pomon
    9      15      28.3      8  1  LCK3_LEUMA             P21142 leucophaea
   10      15      28.3      8  1  NPB_BOVIN              P15507 bos taurus
   11      15      28.3      8  1  UPA1_HUMAN             P30087 homo sapien
   12      14      26.4      6  1  TRPI_PSEPU             P36414 pseudomonas
   13      14      26.4      8  1  GLUR_HUMAN             P02729 homo sapien
   14      13      24.5      6  1  OVM_LEPDE              P42985 leptinotars
   15      13      24.5      7  1  ALL4_CARMA             P81807 carcinus ma
   16      13      24.5      7  1  FAR1_HELTI             P41871 helisoma tr
   17      13      24.5      7  1  FAR4_PANRE             P41875 panagrellus
   18      13      24.5      8  1  ALL6_CYDPO             P82157 cydia pomon
   19      13      24.5      8  1  FAR7_ASCSU             P43171 ascaris suu
   20      12      22.6      6  1  TMOF_SARBU             P41495 sarcophaga
   21      12      22.6      7  1  ALL7_CYDPO             P82158 cydia pomon
   22      12      22.6      8  1  AKH_MELML              P25423 melolontha
   23      12      22.6      8  1  ALL5_CYDPO             P82156 cydia pomon
   24      12      22.6      8  1  LCK4_LEUMA             P21143 leucophaea
   25      12      22.6      8  1  UH09_RAT               P56575 rattus norv
   26      12      22.6      9  1  FAR5_ASCSU             P43170 ascaris suu
   27      12      22.6      9  1  OXYF_SCYCA             P42997 scyliorhinu
   28      12      22.6      9  1  SAP_STOVA              P24047 stomopneute
   29      12      22.6      9  1  UPA7_HUMAN             P30093 homo sapien
   30      11      20.8      4  1  FAR3_HIRME             P42562 hirudo medi
   31      11      20.8      5  1  PRCT_PERAM             P01373 periplaneta
   32      11      20.8      5  1  TRM3_ECOLI             P13973 escherichia
   33      11      20.8      6  1  CIP1_MYTED             P13736 mytilus edu
   34      11      20.8      6  1  CIP2_MYTED             P13737 mytilus edu
   35      11      20.8      7  1  FAR1_ASCSU             P31889 ascaris suu
   36      11      20.8      7  1  FAR2_ASCSU             P31890 ascaris suu
   37      11      20.8      7  1  GFRP_MOUSE             P99025 mus musculu
   38      11      20.8      7  1  MSCL_SALTY             P39446 salmonella
   39      11      20.8      8  1  CAD1_ENTFA             P13268 enterococcu
   40      11      20.8      9  1  DCML_PSECF             P19913 pseudomonas
   41      11      20.8      9  1  NEUU_CAVPO             P34966 cavia porce
   42      11      20.8      9  1  NEUX_HUMAN             P04277 homo sapien
   43      11      20.8      9  1  NEUX_RAT               P11382 rattus norv
   44      11      20.8      9  1  NSK1_SARBU             P41492 sarcophaga
   45      11      20.8      9  1  OXYA_SCYCA             P42996 scyliorhinu
   46      11      20.8      9  1  OXYA_SQUAC             P42999 squalus aca
   47      11      20.8      9  1  UHA2_HUMAN             P40929 homo sapien
   48      10      18.9      5  1  AL14_CARMA             P81817 carcinus ma
   49      10      18.9      7  1  ALL2_CARMA             P81805 carcinus ma
   50      10      18.9      7  1  ALL3_CARMA             P81806 carcinus ma
   51      10      18.9      7  1  ALL5_CARMA             P81808 carcinus ma
   52      10      18.9      7  1  FAR2_PROCL             P38498 procambarus
   53      10      18.9      7  1  UF03_MOUSE             P38641 mus musculu
   54      10      18.9      7  1  UN06_PINPS             P81675 pinus pinas
   55      10      18.9      8  1  AL12_CARMA             P81815 carcinus ma
   56      10      18.9      8  1  AL15_CARMA             P81818 carcinus ma
   57      10      18.9      8  1  AL17_CARMA             P81820 carcinus ma
   58      10      18.9      8  1  AL18_CARMA             P81821 carcinus ma
   59      10      18.9      8  1  ALL3_CYDPO             P82154 cydia pomon
   60      10      18.9      8  1  ALL4_CALVO             P41840 calliphora
   61      10      18.9      8  1  ALL4_CYDPO             P82155 cydia pomon
   62      10      18.9      8  1  ALL5_CALVO             P41841 calliphora
   63      10      18.9      8  1  ALL7_CARMA             P81809 carcinus ma
   64      10      18.9      8  1  ALL8_CARMA             P81811 carcinus ma
   65      10      18.9      8  1  ALL9_CARMA             P81812 carcinus ma
   66      10      18.9      8  1  FAR3_HOMAM             P41486 homarus ame
   67      10      18.9      9  1  AL10_CARMA             P81813 carcinus ma
   68      10      18.9      9  1  AL11_CARMA             P81814 carcinus ma
   69      10      18.9      9  1  FLA2_TREHY             P80159 treponema h
   70      10      18.9      9  1  MOSF_CLYJA             P19853 clypeaster
   71      10      18.9      9  1  OXYT_CYPCA             P23879 cyprinus ca
   72      10      18.9      9  1  OXYT_RABIT             P32878 oryctolagus
   73      10      18.9      9  1  OXYV_SQUAC             P43000 squalus aca
   74      10      18.9      9  1  TAL1_PICJA             P17440 pichia jadi
   75      10      18.9      9  1  TAL3_PICJA             P17441 pichia jadi


                                  ALIGNMENTS


RESULT   1
LCK1_LEUMA
ID   LCK1_LEUMA     STANDARD;      PRT;     8 AA.
AC   P21140;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   01-MAY-1991 (Rel. 18, Last annotation update)
DE   LEUCOKININ I (L-I).
OS   Leucophaea maderae (Madeira cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blaberoidea; Blaberidae; Leucophaea.
RN   [1]
RP   SEQUENCE, AND SYNTHESIS.
RC   TISSUE=HEAD;
RA   Holman G.M., Cook B.J., Nachman R.J.;
RT   "Isolation, primary structure and synthesis of two neuropeptides
RT   from Leucophaea maderae: members of a new family of
RT   Cephalomyotropins.";
RL   Comp. Biochem. Physiol. 84C:205-211(1986).
CC   -!- FUNCTION: THIS CEPHALOMYOTROPIC PEPTIDE STIMULATES CONTRACTILE
CC       ACTIVITY OF COCKROACH PROTODEUM (HINDGUT).
CC   -!- SIMILARITY: TO THE OTHER LEUCOKININS.
KW   Neuropeptide; Amidation.
FT   MOD_RES        8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  893 MW;  DC6365B449CDC76A CRC64;
```

```
  Query Match             45.3%;  Score 24;  DB 1;  Length 8;
  Best Local Similarity   80.0%;  Pred. No. 8.8e+04;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 DPEFN 8
           || ||
Db       1 DPAFN 5


RESULT   2
FAR1_PANRE
ID   FAR1_PANRE     STANDARD;      PRT;     8 AA.
AC   P41872;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE PF1 (SDPNFLRF-AMIDE).
OS   Panagrellus redivivus.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
OC   Panagrolaimoidea; Panagrolaimidae; Panagrellus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 93027659.
RA   Geary T.G., Price D.A., Bowman J.W., Winterrowd C.A., Mackenzie C.D.,
RA   Garrison R.D., Williams J.F., Friedman A.R.;
RT   "Two FMRFamide-like peptides from the free-living nematode
RT   Panagrellus redivivus.";
RL   Peptides 13:209-214(1992).
CC   -!- FUNCTION: MYOACTIVE.
CC   -!- TISSUE SPECIFICITY: NERVE CORDS AND PAIRED GROUPS OF CELLS LOCATED
CC       CAUDALLY TO THE BASE OF THE PHARYNX.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  995 MW;  C6D40729C4576AB5 CRC64;

  Query Match             35.8%;  Score 19;  DB 1;  Length 8;
  Best Local Similarity   75.0%;  Pred. No. 8.8e+04;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 DPEF 7
           || |
Db       2 DPNF 5


RESULT   3
LCK7_LEUMA
ID   LCK7_LEUMA     STANDARD;      PRT;     8 AA.
AC   P19989;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   LEUCOKININ VII (L-VII).
OS   Leucophaea maderae (Madeira cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blaberoidea; Blaberidae; Leucophaea.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=HEAD;
RA   Holman G.M., Cook B.J., Nachman R.J.;
RT   "Isolation, primary structure and synthesis of leucokinins VII and
RT   VIII: the final members of this new family of cephalomyotropic
RT   peptides isolated from head extracts of Leucophaea maderae.";
RL   Comp. Biochem. Physiol. 88C:31-34(1987).
CC   -!- FUNCTION: THIS CEPHALOMYOTROPIC PEPTIDE STIMULATES CONTRACTILE
CC       ACTIVITY OF COCKROACH PROTODEUM (HINDGUT).
CC   -!- SIMILARITY: TO THE OTHER LEUCOKININS.
DR   PIR; JS0317; JS0317.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  866 MW;  DC6365A5B9CDC76A CRC64;

  Query Match             35.8%;  Score 19;  DB 1;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 8.8e+04;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 DPEFN 8
           || |:
Db       1 DPAFS 5


RESULT   4
FAR2_PANRE
ID   FAR2_PANRE     STANDARD;      PRT;     9 AA.
AC   P41873;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE PF2 (SADPNFLRF-AMIDE).
OS   Panagrellus redivivus.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
OC   Panagrolaimoidea; Panagrolaimidae; Panagrellus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 93027659.
RA   Geary T.G., Price D.A., Bowman J.W., Winterrowd C.A., Mackenzie C.D.,
RA   Garrison R.D., Williams J.F., Friedman A.R.;
RT   "Two FMRFamide-like peptides from the free-living nematode
RT   Panagrellus redivivus.";
RL   Peptides 13:209-214(1992).
CC   -!- FUNCTION: MYOACTIVE.
CC   -!- TISSUE SPECIFICITY: NERVE CORDS AND PAIRED GROUPS OF CELLS LOCATED
CC       CAUDALLY TO THE BASE OF THE PHARYNX.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1066 MW;  DA0B0729C4576AAD CRC64;

  Query Match             35.8%;  Score 19;  DB 1;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 8.8e+04;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 DPEF 7
           || |
Db       3 DPNF 6


RESULT   5
LCK2_LEUMA
ID   LCK2_LEUMA     STANDARD;      PRT;     8 AA.
AC   P21141;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   01-MAY-1991 (Rel. 18, Last annotation update)
DE   LEUCOKININ II (L-II).
OS   Leucophaea maderae (Madeira cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blaberoidea; Blaberidae; Leucophaea.
RN   [1]
RP   SEQUENCE, AND SYNTHESIS.
RC   TISSUE=HEAD;
RA   Holman G.M., Cook B.J., Nachman R.J.;
RT   "Isolation, primary structure and synthesis of two neuropeptides
RT   from Leucophaea maderae: members of a new family of
RT   Cephalomyotropins.";
RL   Comp. Biochem. Physiol. 84C:205-211(1986).
CC   -!- FUNCTION: THIS CEPHALOMYOTROPIC PEPTIDE STIMULATES CONTRACTILE
```

```
CC       ACTIVITY OF COCKROACH PROTODEUM (HINDGUT).
CC   -!- SIMILARITY: TO THE OTHER LEUCOKININS.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  852 MW;  DC6365A5B9C8676A CRC64;


  Query Match            34.0%;  Score 18;  DB 1;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 8.8e+04;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 DPEFN 8
           || |:
Db       1 DPGFS 5


RESULT   6
COXE_THUOB
ID   COXE_THUOB     STANDARD;      PRT;     9 AA.
AC   P80975;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE VIA (EC 1.9.3.1) (FRAGMENT).
OS   Thunnus obesus (Bigeye tuna).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;
OC   Scombridae; Thunnus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=HEART;
RX   MEDLINE; 97454291.
RA   Arnold S., Lee I., Kim M., Song E., Linder D., Lottspeich F.,
RA   Kadenbach B.;
RT   "The subunit structure of cytochrome-c oxidase from tuna heart and
RT   liver.";
RL   Eur. J. Biochem. 248:99-103(1997).
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF THE NUCLEAR-CODED POLYPEPTIDE
CC       CHAINS OF CYTOCHROME C OXIDASE, THE TERMINAL OXIDASE IN
CC       MITOCHONDRIAL ELECTRON TRANSPORT.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE VIA FAMILY.
DR   INTERPRO; IPR001349; -.
DR   PROSITE; PS01329; COX6A; PARTIAL.
KW   Oxidoreductase; Inner membrane; Mitochondrion.
FT   NON_TER       1      1
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1136 MW;  62E072C9CB0776DB CRC64;


  Query Match            34.0%;  Score 18;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 PEF 7
           |||
Db       4 PEF 6


RESULT   7
ANG2_BOTJA
ID   ANG2_BOTJA     STANDARD;      PRT;     8 AA.
AC   Q10582;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   ANGIOTENSIN-LIKE PEPTIDE II (FRAGMENT).
OS   Bothrops jararaca (Jararaca).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
```

```
OC   Viperidae; Crotalinae; Bothrops.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PLASMA;
RX   MEDLINE; 96208932.
RA   Borgheresi R.A.M.B., Dalle Lucca J., Carmona E., Picarelli Z.P.;
RT   "Isolation and identification of angiotensin-like peptides from the
RT   plasma of the snake Bothrops jararaca.";
RL   Comp. Biochem. Physiol. 113B:467-473(1996).
CC   -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR   INTERPRO; IPR000215; -.
DR   PROSITE; PS00284; SERPIN; PARTIAL.
KW   Vasoconstrictor; Plasma; Serpin.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  1046 MW;  DDD761E04B42D40A CRC64;


  Query Match            32.1%;  Score 17;  DB 1;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLH 3
           |:|
Db       4 YIH 6


RESULT   8
ALL1_CYDPO
ID   ALL1_CYDPO     STANDARD;      PRT;     8 AA.
AC   P82152;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYDIASTATIN 1.
OS   Cydia pomonella (Codling moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Tortricoidea; Tortricidae; Olethreutinae; Cydia.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LARVA;
RX   MEDLINE; 98054539.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
RA   Davey M., East P.D., Thorpe A.;
RT   "Lepidopteran peptides of the allatostatin superfamily.";
RL   Peptides 18:1301-1309(1997).
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  934 MW;  C82879C45B51F775 CRC64;


  Query Match            30.2%;  Score 16;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 PEFN 8
           | :|
Db       2 PHYN 5


RESULT   9
LCK3_LEUMA
ID   LCK3_LEUMA     STANDARD;      PRT;     8 AA.
AC   P21142;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   01-MAY-1991 (Rel. 18, Last annotation update)
DE   LEUCOKININ III (L-III).
OS   Leucophaea maderae (Madeira cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blaberoidea; Blaberidae; Leucophaea.
```

```
RN   [1]
RP   SEQUENCE, AND SYNTHESIS.
RC   TISSUE=HEAD;
RA   Holman G.M., Cook B.J., Nachman R.J.;
RT   "Primary structure and synthesis of two additional neuropeptides
RT   from Leucophaea maderae: members of a new family of
RT   Cephalomyotropins.";
RL   Comp. Biochem. Physiol. 84C:271-276(1986).
CC   -!- FUNCTION: THIS CEPHALOMYOTROPIC PEPTIDE STIMULATES CONTRACTILE
CC       ACTIVITY OF COCKROACH PROTODEUM (HINDGUT).
CC   -!- SIMILARITY: TO THE OTHER LEUCOKININS.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  910 MW;  DC6365B449C866DA CRC64;

  Query Match             28.3%;  Score 15;  DB 1;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 8.8e+04;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        4 DPEFN 8
            |  ||
Db        1 DQGFN 5

RESULT  10
NPB_BOVIN
ID   NPB_BOVIN      STANDARD;      PRT;     8 AA.
AC   P15507;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-APR-1990 (Rel. 14, Last annotation update)
DE   NEUROPEPTIDE B.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=BRAIN;
RX   MEDLINE; 86067985.
RA   Yang H.-Y.T., Fratta W., Majane E.A., Costa E.;
RT   "Isolation, sequencing, synthesis, and pharmacological
RT   characterization of two brain neuropeptides that modulate the action
RT   of morphine.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:7757-7761(1985).
CC   -!- FUNCTION: MODULATES THE ACTION OF MORPHINE.
DR   PIR; B24749; B24749.
KW   Neuropeptide; Amidation.
FT   SIMILAR       5      8       TO NEUROPEPTIDE A (AA 5-8) (IDENTICAL).
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  1082 MW;  87D416C776D9C729 CRC64;

  Query Match             28.3%;  Score 15;  DB 1;  Length 8;
  Best Local Similarity   33.3%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLHDPE 6
             :|  |:
Db        1 FLFQPQ 6

RESULT  11
UPA1_HUMAN
ID   UPA1_HUMAN     STANDARD;      PRT;     8 AA.
AC   P30087;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE OF PLASMA (SPOT 2) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PLASMA;
RX   MEDLINE; 93092937.
RA   Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,
RA   Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,
RA   Hochstrasser D.F.;
RT   "Plasma protein map: an update by microsequencing.";
RL   Electrophoresis 13:707-714(1992).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 4.9, ITS MW IS: 65 KDA.
DR   SWISS-2DPAGE; P30087; HUMAN.
FT   NON_TER       1      1
FT   UNSURE        8      8
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  944 MW;  C01772C455BB06DA CRC64;

  Query Match             28.3%;  Score 15;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        4 DPEFNL 9
            | | |:
Db        1 DQESNV 6

RESULT  12
TRPI_PSEPU
ID   TRPI_PSEPU     STANDARD;      PRT;     6 AA.
AC   P36414;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   TRPBA OPERON TRANSCRIPTIONAL ACTIVATOR (FRAGMENT).
GN   TRPI.
OS   Pseudomonas putida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PPG1 C1S;
RX   MEDLINE; 89335826.
RA   Eberly L., Crawford I.P.;
RT   "DNA sequence of the tryptophan synthase genes of Pseudomonas
RT   putida.";
RL   Biochimie 71:521-531(1989).
CC   -!- FUNCTION: ACTIVATES THE EXPRESSION OF THE TRPBA GENES ENCODING
CC       THE TWO TRYPTOPHAN SYNTHASE SUBUNITS. IN THE ABSENCE OF THE
CC       INDUCER (INDOLEGLYCEROL PHOSPHATE), TRPI BINDS UPSTREAM OF THE
CC       TRPAB OPERON, OVERLAPPING ITS OWN PROMOTER REGION.
CC   -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC       REGULATORS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X13299; CAA31660.1; -.
DR   INTERPRO; IPR000847; -.
DR   PROSITE; PS00044; HTH_LYSR_FAMILY; PARTIAL.
KW   Tryptophan biosynthesis; Transcription regulation; Activator;
KW   DNA-binding.
FT   NON_TER       6      6
SQ   SEQUENCE   6 AA;  683 MW;  77672AA1EDD6F000 CRC64;

  Query Match             26.4%;  Score 14;  DB 1;  Length 6;
```

```
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 HD 4
           ||
Db       3 HD 4


RESULT  13
GLUR_HUMAN
ID   GLUR_HUMAN     STANDARD;      PRT;     8 AA.
AC   P02729;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   URINE GLYCOPEPTIDE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 72062338.
RA   Lote C.J., Weiss J.B.;
RT   "Identification in urine of a low-molecular-weight highly polar
RT   glycopeptide containing cysteinyl-galactose.";
RL   Biochem. J. 123:25P-25P(1971).
CC   -!- FUNCTION: THE IDENTITY OF THE GLYCOPROTEIN FROM WHICH THIS PEPTIDE
CC       IS DERIVED IS UNKNOWN. NO PHYSIOLOGICAL FUNCTION HAS BEEN
CC       ATTRIBUTED. AN ERYTHROCYTE MEMBRANE GLYCOPEPTIDE HAVING A
CC       SIMILAR STRUCTURE HAS ALSO BEEN FOUND.
DR   PIR; A03188; XGHUEU.
KW   Glycoprotein.
FT   CARBOHYD      1      1       S-LINKED (GAL. . .).
SQ   SEQUENCE   8 AA;  855 MW;  C2D87AA1F5B1EB1E CRC64;

  Query Match             26.4%;  Score 14;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 HD 4
           ||
Db       5 HD 6


RESULT  14
OVM_LEPDE
ID   OVM_LEPDE      STANDARD;      PRT;     6 AA.
AC   P42985;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   OVIDUCTAL MOTILITY STIMULATING PEPTIDE (LED-OVM).
OS   Leptinotarsa decemlineata (Colorado potato beetle).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC   Cucujiformia; Chrysomelidae; Chrysomelinae; Leptinotarsa.
RN   [1]
RP   SEQUENCE, AND SYNTHESIS.
RC   TISSUE=HEAD;
RX   MEDLINE; 91271080.
RA   Spittaels K., Schoofs L., Grauwels L., Smet H., van Damme J.,
RA   Proost P., Torrekens S., de Loof A.;
RT   "Isolation, identification and synthesis of novel oviductal motility
RT   stimulating head peptide in the Colorado potato beetle, Leptinotarsa
RT   decemlineata.";
RL   Peptides 12:31-36(1991).
CC   -!- FUNCTION: MYOTROPIC PEPTIDE. STIMULATES THE CONTRACTIONS OF THE
CC       OVIDUCT.
KW   Neuropeptide; Amidation.
FT   MOD_RES       6      6       AMIDATION.
SQ   SEQUENCE   6 AA;  720 MW;  6B07632B5DD03000 CRC64;
```

```
  Query Match             24.5%;  Score 13;  DB 1;  Length 6;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 HDPE 6
           : ||
Db       3 YKPE 6


RESULT  15
ALL4_CARMA
ID   ALL4_CARMA     STANDARD;      PRT;     7 AA.
AC   P81807;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 4.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Multigene family.
SQ   SEQUENCE   7 AA;  782 MW;  672879CDCB476AC0 CRC64;

  Query Match             24.5%;  Score 13;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 DP 5
           ||
Db       1 DP 2


RESULT  16
FAR1_HELTI
ID   FAR1_HELTI     STANDARD;      PRT;     7 AA.
AC   P41871;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE GDPFLRF-AMIDE.
OS   Helisoma trivolvis (Snail).
OC   Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;
OC   Planorbidae; Helisoma.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=KIDNEY;
RX   MEDLINE; 94286417.
RA   Madrid K.P., Price D.A., Greenberg M.J., Khan H.R., Saleuddin A.S.M.;
RT   "FMRFamide-related peptides from the kidney of the snail, Helisoma
RT   trivolvis.";
RL   Peptides 15:31-36(1994).
CC   -!- FUNCTION: APPEARS TO BE INVOLVED IN OSMOREGULATION BY AFFECTING
CC       THE KIDNEY, MANTLE AND SKIN.
CC   -!- TISSUE SPECIFICITY: KIDNEY, SKIN, MANTLE AND THE HEMOLYMPH.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  851 MW;  69D40729D76AA810 CRC64;
```

```
  Query Match             24.5%;  Score 13;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 DP 5
           ||
Db       2 DP 3

RESULT  17
FAR4_PANRE
ID   FAR4_PANRE     STANDARD;      PRT;     7 AA.
AC   P41875;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE PF4 (KPNFIRFAMIDE).
OS   Panagrellus redivivus.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
OC   Panagrolaimoidea; Panagrolaimidae; Panagrellus.
RN   [1]
RP   SEQUENCE, AND SYNTHESIS.
RX   MEDLINE; 95232026.
RA   Maule A.G., Shaw C., Bowman J.W., Halton D.W., Thompson D.P.,
RA   Thim L., Kubiak T.M., Martin R.A., Geary T.G.;
RT   "Isolation and preliminary biological characterization of
RT   KPNFIRFamide, a novel FMRFamide-related peptide from the free-living
RT   nematode, Panagrellus redivivus.";
RL   Peptides 16:87-93(1995).
CC   -!- FUNCTION: MYOACTIVE; INDUCES A RAPID CONCENTRATION-DEPENDENT
CC       MUSCLE TENSION INCREASE.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  921 MW;  69D40059C4576350 CRC64;

  Query Match             24.5%;  Score 13;  DB 1;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 PEF 7
           | |
Db       2 PNF 4

RESULT  18
ALL6_CYDPO
ID   ALL6_CYDPO     STANDARD;      PRT;     8 AA.
AC   P82157;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYDIASTATIN 6.
OS   Cydia pomonella (Codling moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Tortricoidea; Tortricidae; Olethreutinae; Cydia.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LARVA;
RX   MEDLINE; 98054539.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
RA   Davey M., East P.D., Thorpe A.;
RT   "Lepidopteran peptides of the allatostatin superfamily.";
RL   Peptides 18:1301-1309(1997).
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  936 MW;  0B2879C45B573767 CRC64;
```

```
  Query Match             24.5%;  Score 13;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 PEFN 8
           | :|
Db       2 PLYN 5

RESULT  19
FAR7_ASCSU
ID   FAR7_ASCSU     STANDARD;      PRT;     8 AA.
AC   P43171;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE AF7.
OS   Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC   Ascarididae; Ascaris.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 95380362.
RA   Cowden C., Stretton A.O.W.;
RT   "Eight novel FMRFamide-like neuropeptides isolated from the nematode
RT   Ascaris suum.";
RL   Peptides 16:491-500(1995).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  963 MW;  9CD40059D417687D CRC64;

  Query Match             24.5%;  Score 13;  DB 1;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 PEF 7
           | |
Db       3 PRF 5

RESULT  20
TMOF_SARBU
ID   TMOF_SARBU     STANDARD;      PRT;     6 AA.
AC   P41495;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   TRYPSIN-MODULATING OOSTATIC FACTOR (TMOF).
OS   Sarcophaga bullata (Grey flesh fly) (Neobellieria bullata).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Sarcophagidae; Sarcophaga.
RN   [1]
RP   SEQUENCE, AND SYNTHESIS.
RC   TISSUE=OVARY;
RX   MEDLINE; 94211930.
RA   Bylemans D., Borovsky D., Hunt D.F., Shabanowitz J., Grauwels L.,
RA   de Loof A.;
RT   "Sequencing and characterization of trypsin modulating oostatic
RT   factor (TMOF) from the ovaries of the grey fleshfly, Neobellieria
RT   (Sarcophaga) bullata.";
RL   Regul. Pept. 50:61-72(1994).
CC   -!- FUNCTION: HAS AN OOSTATIC ACTIVITY. INHIBITS TRYPSIN BIOSYNTHESIS
CC       IN THE MIDGUT WHICH INDIRECTLY REDUCES THE VITELLOGENIN
CC       CONCENTRATION IN THE HEMOLYMPH RESULTING IN INHIBITION OF OOCYTE
CC       DEVELOPMENT.
CC   -!- DEVELOPMENTAL STAGE: SYNTHESIZED AND RELEASED FROM FOLLICULAR
CC       EPITHELIUM AFTER A BLOOD MEAL.
```

```
KW   Hormone.
SQ   SEQUENCE   6 AA;  695 MW;  61E72451B7642000 CRC64;

  Query Match             22.6%;  Score 12;  DB 1;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LH 3
           ||
Db       5 LH 6


RESULT  21
ALL7_CYDPO
ID   ALL7_CYDPO     STANDARD;      PRT;     7 AA.
AC   P82158;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYDIASTATIN 7.
OS   Cydia pomonella (Codling moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Tortricoidea; Tortricidae; Olethreutinae; Cydia.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LARVA;
RX   MEDLINE; 98054539.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
RA   Davey M., East P.D., Thorpe A.;
RT   "Lepidopteran peptides of the allatostatin superfamily.";
RL   Peptides 18:1301-1309(1997).
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  873 MW;  672879CABB569350 CRC64;

  Query Match             22.6%;  Score 12;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       6 EFNL 9
           :| |
Db       4 DFGL 7


RESULT  22
AKH_MELML
ID   AKH_MELML      STANDARD;      PRT;     8 AA.
AC   P25423;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   ADIPOKINETIC HORMONE (AKH).
OS   Melolontha melolontha (Cockchafer),
OS   Geotrupes stercorosus (Dor beetle), and
OS   Pachnoda marginata (Flower beetle).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC   Scarabaeiformia; Scarabaeidae; Melolonthinae; Melolontha.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=M.MELOLONTHA, AND G.STERCOROSUS; TISSUE=CORPORA CARDIACA;
RX   MEDLINE; 91248100.
RA   Gaede G.;
RT   "A unique charged tyrosine-containing member of the adipokinetic
RT   hormone/red-pigment-concentrating hormone peptide family isolated and
RT   sequenced from two beetle species.";
RL   Biochem. J. 275:671-677(1991).
RN   [2]
RP   SEQUENCE.
RC   SPECIES=P.MARGINATA; TISSUE=CORPORA CARDIACA;
RX   MEDLINE; 92265187.
RA   Gaede G., Lopata A., Kellner R., Rinehart K.L. Jr.;
RT   "Primary structures of neuropeptides isolated from the corpora
RT   cardiaca of various cetonid beetle species determined by
RT   pulsed-liquid phase sequencing and tandem fast atom bombardment mass
RT   spectrometry.";
RL   Biol. Chem. Hoppe-Seyler 373:133-142(1992).
CC   -!- FUNCTION: THIS HORMONE, RELEASED FROM CELLS IN THE CORPORA
CC       CARDIACA AFTER THE BEGINNING OF FLIGHT, CAUSES RELEASE OF
CC       DIGLYCERIDES FROM THE FAT BODY AND THEN STIMULATES THE FLIGHT
CC       MUSCLES TO USE THESE DIGLYCERIDES AS AN ENERGY SOURCE.
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; S15422; S15422.
DR   PIR; S21663; S21663.
DR   INTERPRO; IPR002047; -.
DR   PROSITE; PS00256; AKH; 1.
KW   Neuropeptide; Amidation; Flight.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  1022 MW;  867AB775AB544736 CRC64;

  Query Match             22.6%;  Score 12;  DB 1;  Length 8;
  Best Local Similarity   20.0%;  Pred. No. 8.8e+04;
  Matches    1;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 HDPEF 7
           : |::
Db       4 YSPDW 8


RESULT  23
ALL5_CYDPO
ID   ALL5_CYDPO     STANDARD;      PRT;     8 AA.
AC   P82156;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYDIASTATIN 5.
OS   Cydia pomonella (Codling moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Tortricoidea; Tortricidae; Olethreutinae; Cydia.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LARVA;
RX   MEDLINE; 98054539.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
RA   Davey M., East P.D., Thorpe A.;
RT   "Lepidopteran peptides of the allatostatin superfamily.";
RL   Peptides 18:1301-1309(1997).
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  898 MW;  922879CABB58640D CRC64;

  Query Match             22.6%;  Score 12;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       6 EFNL 9
           :| |
Db       5 DFGL 8


RESULT  24
LCK4_LEUMA
ID   LCK4_LEUMA     STANDARD;      PRT;     8 AA.
AC   P21143;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
```

```
DT   01-MAY-1991 (Rel. 18, Last annotation update)
DE   LEUCOKININ IV (L-IV).
OS   Leucophaea maderae (Madeira cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blaberoidea; Blaberidae; Leucophaea.
RN   [1]
RP   SEQUENCE, AND SYNTHESIS.
RC   TISSUE=HEAD;
RA   Holman G.M., Cook B.J., Nachman R.J.;
RT   "Primary structure and synthesis of two additional neuropeptides
RT   from Leucophaea maderae: members of a new family of
RT   Cephalomyotropins.";
RL   Comp. Biochem. Physiol. 84C:271-276(1986).
CC   -!- FUNCTION: THIS CEPHALOMYOTROPIC PEPTIDE STIMULATES CONTRACTILE
CC       ACTIVITY OF COCKROACH PROTODEUM (HINDGUT).
CC   -!- SIMILARITY: TO THE OTHER LEUCOKININS.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  906 MW;  DC6365B1E9D5BDDA CRC64;


  Query Match             22.6%;  Score 12;  DB 1;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 DPEFN 8
           |  |:
Db       1 DASFH 5


RESULT  25
UH09_RAT
ID   UH09_RAT       STANDARD;      PRT;     8 AA.
AC   P56575;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE OF HEART TISSUE (SPOT P9) (FRAGMENT).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE.
RC   STRAIN=WISTAR; TISSUE=HEART;
RA   Li X.-P., Pleissner K.-P., Scheler C., Regitz-Zagrosek V., Salikov J.,
RA   Jungblut P.R.;
RL   Submitted (SEP-1998) to the SWISS-PROT data bank.
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 8.9, ITS MW IS: 42 KDA.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  1029 MW;  9E0775A6C4140B06 CRC64;


  Query Match             22.6%;  Score 12;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 PE 6
           ||
Db       7 PE 8


RESULT  26
FAR5_ASCSU
ID   FAR5_ASCSU     STANDARD;      PRT;     9 AA.
AC   P43170;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE AF5.
OS   Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
```

```
OC   Ascarididae; Ascaris.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 95380362.
RA   Cowden C., Stretton A.O.W.;
RT   "Eight novel FMRFamide-like neuropeptides isolated from the nematode
RT   Ascaris suum.";
RL   Peptides 16:491-500(1995).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1052 MW;  340B0059D1B76338 CRC64;


  Query Match             22.6%;  Score 12;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 PEF 7
           | |
Db       4 PTF 6


RESULT  27
OXYF_SCYCA
ID   OXYF_SCYCA     STANDARD;      PRT;     9 AA.
AC   P42997;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   PHASVATOCIN.
OS   Scyliorhinus canicula (Spotted dogfish) (Spotted catshark).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Galeomorphii; Galeoidea; Carcharhiniformes;
OC   Scyliorhinidae; Scyliorhinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PITUITARY;
RX   MEDLINE; 95062247.
RA   Chauvet J., Rouille Y., Chauveau C., Chauvet M.-T., Acher R.;
RT   "Special evolution of neurohypophysial hormones in cartilaginous
RT   fishes: asvatocin and phasvatocin, two oxytocin-like peptides
RT   isolated from the spotted dogfish (Scyliorhinus caniculus).";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:11266-11270(1994).
CC   -!- FUNCTION: DISPLAYS OXYTOCIC ACTIVITY ON RAT UTERUS.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1016 MW;  17EDD76EB44449DB CRC64;


  Query Match             22.6%;  Score 12;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 FN 8
           ||
Db       3 FN 4


RESULT  28
SAP_STOVA
ID   SAP_STOVA      STANDARD;      PRT;     9 AA.
AC   P24047;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   01-MAR-1992 (Rel. 21, Last annotation update)
DE   SPERM-ACTIVATING PEPTIDE (SAP).
```

```
OS   Stomopneutes variolaris (Sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Diadematacea; Phymosomatoida; Stomechinidae;
OC   Stomopneustes.
RN   [1]
RP   SEQUENCE, AND DISULFIDE BOND.
RC   TISSUE=EGG JELLY;
RX   MEDLINE; 92097763.
RA   Yoshimo K.-I., Takao T., Shimonishi Y., Suzuki N.;
RT   "Determination of the amino acid sequence of an intramolecular
RT   disulfide linkage-containing sperm-activating peptide by tandem mass
RT   spectrometry.";
RL   FEBS Lett. 294:179-182(1991).
CC   -!- FUNCTION: CAUSE STIMULATION OF SPERM RESPIRATION AND MOTILITY
CC       THROUGH INTRACELLULAR ALKALINIZATION, TRANSIENT ELEVATIONS OF
CC       CAMP, CGMP AND CLACIUM LEVELS IN SPERM CELLS, AND TRANSIENT
CC       ACTIVATION AND SUBSEQUENT INACTIVATION OF THE MEMBRANE FORM OF
CC       GUANYLATE CYCLASE.
DR   PIR; S19329; S19329.
FT   DISULFID      3      8
SQ   SEQUENCE   9 AA;  1010 MW;  C469B3387B076EB9 CRC64;

  Query Match             22.6%;  Score 12;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 PE 6
           ||
Db       4 PE 5


RESULT  29
UPA7_HUMAN
ID   UPA7_HUMAN     STANDARD;      PRT;     9 AA.
AC   P30093;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE OF PLASMA (SPOT 18) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PLASMA;
RX   MEDLINE; 93092937.
RA   Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,
RA   Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,
RA   Hochstrasser D.F.;
RT   "Plasma protein map: an update by microsequencing.";
RL   Electrophoresis 13:707-714(1992).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL`THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 5.05, ITS MW IS: 37 KDA.
DR   SWISS-2DPAGE; P30093; HUMAN.
FT   NON_TER       1      1
FT   UNSURE        5      5
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1042 MW;  5C14477AEB0772C7 CRC64;

  Query Match             22.6%;  Score 12;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 PE 6
           ||
Db       4 PE 5


RESULT  30
FAR3_HIRME
ID   FAR3_HIRME     STANDARD;      PRT;     4 AA.
AC   P42562;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE YLRF-AMIDE.
OS   Hirudo medicinalis (Medicinal leech).
OC   Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;
OC   Arynchobdellida; Hirudiniformes; Hirudinidae; Hirudo.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92195954.
RA   Evans B.D., Pohl J., Kartsonis M.A., Calabrese R.L.;
RT   "Identification of RFamide neuropeptides in the medicinal leech.";
RL   Peptides 12:897-908(1991).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       4      4       AMIDATION.
SQ   SEQUENCE   4 AA;  598 MW;  69D4073B30000000 CRC64;

  Query Match             20.8%;  Score 11;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YL 2
           ||
Db       1 YL 2


RESULT  31
PRCT_PERAM
ID   PRCT_PERAM     STANDARD;      PRT;     5 AA.
AC   P01373;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   PROCTOLIN.
OS   Periplaneta americana (American cockroach),
OS   Limulus polyphemus (Atlantic horseshoe crab), and
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blattoidea; Blattidae; Periplaneta.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=P.AMERICANA;
RX   MEDLINE; 76074708.
RA   Starratt A.N., Brown B.E.;
RT   "Structure of the pentapeptide proctolin, a proposed neurotransmitter
RT   in insects.";
RL   Life Sci. 17:1253-1256(1975).
RN   [2]
RP   BIOLOGICAL SOURCE.
RC   SPECIES=P.AMERICANA;
RX   MEDLINE; 81225865.
RA   O'Shea M., Adams M.E.;
RT   "Pentapeptide (proctolin) associated with an identified neuron.";
RL   Science 213:567-569(1981).
RN   [3]
RP   SEQUENCE.
RC   SPECIES=L.POLYPHEMUS;
RX   MEDLINE; 90287800.
RA   Groome J.R., Tillinghast E.K., Townley M.A., Vetrovs A.,
RA   Watson W.H. III, Hunt D.F., Griffin P.R., Alexander J.E.,
RA   Shabanowitz J.;
RT   "Identification of proctolin in the central nervous system of the
RT   horseshoe crab, Limulus polyphemus.";
RL   Peptides 11:205-211(1990).
RN   [4]
RP   SEQUENCE.
RC   SPECIES=C.MAENAS;
RX   MEDLINE; 86232789.
```

```
RA   Stangier J., Dircksen H., Keller R.;
RT   "Identification and immunocytochemical localization of proctolin in
RT   pericardial organs of the shore crab, Carcinus maenas.";
RL   Peptides 7:67-72(1986).
CC   -!- FUNCTION: STIMULATES CARDIAC OUTPUT AND HINDGUT MOTILITY,
CC       MODULATES VISCERAL AND SKELETAL MUSCLE IN MANY ARTHROPODS.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE LATERAL WHITE NEURONS AND IN
CC       THE CRAB PERICARDIAL ORGANS.
DR   PIR; A01644; HOROHA.
DR   PIR; A60411; A60411.
KW   Neuropeptide.
SQ   SEQUENCE   5 AA;  649 MW;  71B7673B44600000 CRC64;


  Query Match             20.8%;  Score 11;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YL 2
            ||
Db        2 YL 3


RESULT  32
TRM3_ECOLI
ID   TRM3_ECOLI     STANDARD;      PRT;     5 AA.
AC   P13973;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   TRAM PROTEIN (FRAGMENT).
GN   TRAM.
OS   Escherichia coli.
OG   Plasmid IncFII R100.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88227859.
RA   Inamoto S., Yoshioka Y., Ohtsubo E.;
RT   "Identification and characterization of the products from the traJ
RT   and traY genes of plasmid R100.";
RL   J. Bacteriol. 170:2749-2757(1988).
CC   -!- FUNCTION: TRANSFER GENE PROTEIN. IS INVOLVED IN THE CONJUGATION
CC       PROCESS OF BACTERIAL CELLS FOR THE EXCHANGE OF PLASMID DNA.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: TO TRAM PROTEIN OF OTHER PLASMIDS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M20941; -; NOT_ANNOTATED_CDS.
DR   PIR; A32014; A32014.
KW   Conjugation; Plasmid; DNA-binding.
FT   NON_TER       1      1
SQ   SEQUENCE   5 AA;  634 MW;  6B1B1AA443500000 CRC64;


  Query Match             20.8%;  Score 11;  DB 1;  Length 5;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        3 HDPE 6
            :| |
Db        2 NDEE 5


RESULT  33
CIP1_MYTED
ID   CIP1_MYTED     STANDARD;      PRT;     6 AA.
AC   P13736;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   CONTRACTION-INHIBITING PEPTIDE I (MIP I).
OS   Mytilus edulis (Blue mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytiloidea; Mytilidae; Mytilus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PEDAL GANGLION;
RX   MEDLINE; 88240357.
RA   Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.;
RT   "Structures and actions of Mytilus inhibitory peptides.";
RL   Biochem. Biophys. Res. Commun. 152:1376-1382(1988).
CC   -!- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN
CC       MUSCLES.
CC   -!- SIMILARITY: TO MIP II.
DR   PIR; A27696; A27696.
KW   Hormone; Amidation.
FT   MOD_RES       6      6       AMIDATION.
SQ   SEQUENCE   6 AA;  637 MW;  72C9C68775B81000 CRC64;


  Query Match             20.8%;  Score 11;  DB 1;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        5 PEF 7
            | |
Db        3 PMF 5


RESULT  34
CIP2_MYTED
ID   CIP2_MYTED     STANDARD;      PRT;     6 AA.
AC   P13737;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   CONTRACTION-INHIBITING PEPTIDE II (MIP II).
OS   Mytilus edulis (Blue mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytiloidea; Mytilidae; Mytilus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PEDAL GANGLION;
RX   MEDLINE; 88240357.
RA   Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.;
RT   "Structures and actions of Mytilus inhibitory peptides.";
RL   Biochem. Biophys. Res. Commun. 152:1376-1382(1988).
CC   -!- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN
CC       MUSCLES.
CC   -!- SIMILARITY: TO MIP I.
DR   PIR; B27696; B27696.
KW   Hormone; Amidation.
FT   MOD_RES       6      6       AMIDATION.
SQ   SEQUENCE   6 AA;  621 MW;  72C9C6876DD81000 CRC64;


  Query Match             20.8%;  Score 11;  DB 1;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        5 PEF 7
            | |
Db        3 PMF 5


RESULT  35
FAR1_ASCSU
```

```
ID   FAR1_ASCSU     STANDARD;      PRT;     7 AA.
AC   P31889;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE AF1.
OS   Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC   Ascarididae; Ascaris.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 90180465.
RA   Cowden C., Stretton A.O.W., Davis R.E.;
RT   "AF1, a sequenced bioactive neuropeptide isolated from the nematode
RT   Ascaris suum.";
RL   Neuron 2:1465-1473(1989).
CC   -!- FUNCTION: POTENT MODULATOR OF INHIBITORY MOTORNEURONS. REDUCES THE
CC       INPUT RESISTANCE AND BLOCKS SLOW OSCILLATORY POTENTIALS IN THESE
CC       CELLS.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE NERVE CORDS AND A VARIETY OF
CC       GANGLIA PARTICULARLY IN THE ANTERIOR REGIONS.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  953 MW;  69D40059CB144350 CRC64;


  Query Match             20.8%;  Score 11;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 EF 7
            ||
Db        3 EF 4


Search completed: December 16, 2000, 04:23:32
Job time: 4567 sec
```

```
                        GenCore version 4.5
              Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 03:35:13 ; Search time 89.11 Seconds
                                              (without alignments)
                                              6.409 Million cell updates/sec

Title:          US-09-529-121-10
Perfect score:  53
Sequence:       1 YLHDPEFNL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters:      793

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :      PIR_65:*
               1:  pir1:*
               2:  pir2:*
               3:  pir3:*
               4:  pir4:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                                  SUMMARIES
                  %
Result          Query
  No.   Score   Match Length DB  ID                    Description
-------------------------------------------------------------------------
    1      20    37.7      7  2  I46868                alpha-myosin heavy
    2      19    35.8      8  2  JS0317                leucokinin VII - M
    3      18    34.0      6  2  A19780                transferrin - bovi
    4      18    34.0      8  2  I64832                Ca2+-transporting
    5      18    34.0      9  2  S77984                cytochrome-c oxida
    6      15    28.3      6  2  PT0709                T-cell receptor be
    7      15    28.3      8  2  B24749                neuropeptide B - b
    8      15    28.3      9  2  D58503                translation elonga
    9      15    28.3      9  2  A60108                exotoxin A - Strep
   10      14    26.4      8  2  XGHUEU                urine glycopeptide
   11      14    26.4      8  2  A28719                thymic humoral fac
   12      14    26.4      9  2  PT0231                Ig heavy chain CDR
   13      13    24.5      4  2  I51049                metallothionein-A
   14      13    24.5      5  2  B60274                major protein anti
   15      13    24.5      7  2  PT0665                T-cell receptor be
   16      13    24.5      8  2  S71919                alcohol dehydrogen
   17      13    24.5      8  2  PT0368                Ig gamma chain C r
   18      13    24.5      9  2  B45796                dihydrolipoamide S
   19      13    24.5      9  2  G58502                kidney and bladder
   20      13    24.5      9  2  S36898                ribosomal protein
   21      13    24.5      9  2  PT0080                60K Ca binding pro
   22      13    24.5      9  2  G56978                collagen alpha 1(I
   23      13    24.5      9  2  PD0443                3-oxoacid CoA-tran
   24      13    24.5      9  2  S39437                D-amino-acid oxida
   25      12    22.6      4  2  S53508                starvation-induced
   26      12    22.6      4  2  I54357                schwannomin - mous
   27      12    22.6      5  2  A44692                fulicin - giant Af
   28      12    22.6      6  2  JN0861                peptidyl-dipeptida
   29      12    22.6      6  2  S11556                hydrogensulfite re
   30      12    22.6      6  2  S78764                ribosomal protein
   31      12    22.6      7  2  PQ0663                membrane protein -
   32      12    22.6      7  2  H33098                180K exoantigen -
   33      12    22.6      7  2  E48394                glycoprotein compo
   34      12    22.6      7  2  B39040                calsequestrin, fas
   35      12    22.6      7  2  I48293                transforming prote
   36      12    22.6      7  2  PH0932                T-cell receptor be
   37      12    22.6      7  2  B48394                major fat-globule
   38      12    22.6      7  2  S45648                Na+-transporting A
   39      12    22.6      8  2  A32523                peptidyl-dipeptida
   40      12    22.6      8  2  S15422                adipokinetic hormo
   41      12    22.6      8  2  A58641                adipokinetic hormo
   42      12    22.6      8  2  S21273                cellulase (EC 3.2.
   43      12    22.6      8  2  S16324                hypothetical prote
   44      12    22.6      8  2  PT0030                inulinase (EC 3.2.
   45      12    22.6      8  2  S21663                neuropeptide - flo
   46      12    22.6      8  2  S29272                tocopherol-binding
   47      12    22.6      8  2  I57532                gene TnIslow prote
   48      12    22.6      9  2  S02384                probable membrane
   49      12    22.6      9  2  S66607                quinoline 2-oxidor
   50      12    22.6      9  2  T31612                hypothetical prote
   51      12    22.6      9  2  S65433                bradykinin - horn
   52      12    22.6      9  2  S19329                sperm-activating p
   53      12    22.6      9  2  A61363                bradykinin - commo
   54      12    22.6      9  2  I70040                glutathione transf
   55      12    22.6      9  2  S15407                MHC ligand protein
   56      12    22.6      9  2  A43065                hydroxyproline-3-b
   57      11    20.8      5  1  HOROHA                proctolin - Americ
   58      11    20.8      5  2  A32014                traM protein - Esc
   59      11    20.8      5  2  A60411                proctolin - Atlant
   60      11    20.8      6  2  B44510                hypothetical prote
   61      11    20.8      6  2  A27696                contraction-inhibi
   62      11    20.8      6  2  B27696                contraction-inhibi
   63      11    20.8      7  2  S78024                ribosomal protein
   64      11    20.8      8  2  T13818                cytochrome oxidase
   65      11    20.8      8  2  PT0311                Ig heavy chain CRD
   66      11    20.8      8  2  B45800                serum albumin - do
   67      11    20.8      9  2  A61230                calsequestrin, car
   68      11    20.8      9  2  A61057                Thr-6 bradykinin -
   69      11    20.8      9  2  PL0139                carbon-monoxide de
   70      11    20.8      9  2  S66419                tetrameric protein
   71      11    20.8      9  2  A26744                bradykinin-like pe
   72      11    20.8      9  2  S10920                venom protein HR-3
   73      11    20.8      9  2  A61358                bradykinin-like pe
   74      11    20.8      9  2  A60579                bradykinin-like pe
   75      11    20.8      9  2  B60246                ornitho-kinin - ch


                                 ALIGNMENTS


RESULT   1
I46868
alpha-myosin heavy chain - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 05-Nov-1999
C;Accession: I46868
R;Friedman, D.J.; Umeda, P.K.; Sinha, A.M.; Hsu, H.
Proc. Natl. Acad. Sci. U.S.A. 81, 3044-3048, 1984
A;Title: Characterization of genomic clones specifying rabbit alpha- and beta-ventric
A;Reference number: I46868; MUID:84221901
A;Accession: I46868
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-7 <FRI>
A;Cross-references: GB:K01698; NID:g165538; PIDN:AAA31415.1; PID:g165539


  Query Match             37.7%;  Score 20;  DB 2;  Length 7;
  Best Local Similarity   60.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LHDPE 6
```

```
          :|| |
Db      3 MHDEE 7


RESULT   2
JS0317
leucokinin VII - Madeira cockroach
C;Species: Leucophaea maderae (Madeira cockroach)
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 20-Jun-2000
C;Accession: JS0317
R;Holman, G.M.; Cook, B.J.; Nachman, R.J.
Comp. Biochem. Physiol. C 88, 31-34, 1987
A;Title: Isolation, primary structure and synthesis of leucokinins VII and VIII: the fir
A;Reference number: JS0317
A;Accession: JS0317
A;Molecule type: protein
A;Residues: 1-8 <HOL>
C;Comment: Leucokinins, a family of cephalomyotropic peptides, stimulate contractile act
C;Keywords: amidated carboxyl end; cephalomyotropic peptide
F;8/Modified site: amidated carboxyl end (Gly) #status experimental


  Query Match             35.8%;  Score 19;  DB 2;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 DPEFN 8
          || |:
Db      1 DPAFS 5


RESULT   3
A19780
transferrin - bovine (fragment)
C;Species: Bos primigenius taurus (cattle)
C;Date: 25-Oct-1987 #sequence_revision 25-Oct-1987 #text_change 18-Jun-1993
C;Accession: A19780
R;Brock, J.H.; Esparza, I.; Oliver, R.A.; Spooner, R.L.
Biochem. Genet. 18, 851-860, 1980
A;Title: Electrophoretic mobility of N- and C-terminal monoferric fragments of bovine tr
A;Reference number: A19780; MUID:81183891
A;Accession: A19780
A;Molecule type: protein
A;Residues: 1-6 <BRO>


  Query Match             34.0%;  Score 18;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 DPE 6
          |||
Db      1 DPE 3


RESULT   4
I64832
Ca2+-transporting ATPase (EC 3.6.1.38) - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C;Accession: I64832
R;Wu, K.
Am. J. Physiol. 264, 333-341, 1993
A;Title: Molecular cloning and quantification of mRNA and protein encoding Ca2+-ATPase i
A;Reference number: I51892
A;Accession: I64832
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-8 <RES>
A;Cross-references: GB:M99223; NID:g203644; PIDN:AAA40992.1; PID:g203646
C;Genetics:
A;Gene: SERCA1b
C;Keywords: hydrolase
```

```
  Query Match             34.0%;  Score 18;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 DPE 6
          |||
Db      1 DPE 3


RESULT   5
S77984
cytochrome-c oxidase (EC 1.9.3.1) chain VIa - bigeye tuna (fragment)
C;Species: Thunnus obesus (bigeye tuna)
C;Date: 17-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 30-Jan-1998
C;Accession: S77984
R;Arnold, S.; Lee, J.; Kim, M.; Song, E.; Linder, D.; Lottspeich, F.; Kadenbach, B.
submitted to the Protein Sequence Database, June 1997
A;Reference number: S77980
A;Accession: S77984
A;Molecule type: protein
A;Residues: 1-9 <ARN>
A;Experimental source: heart
C;Genetics:
A;Genome: nuclear
C;Function:
A;Pathway: oxidative phosphorylation; respiratory chain
C;Keywords: electron transfer; membrane-associated complex; mitochondrial inner membr


  Query Match             34.0%;  Score 18;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 PEF 7
          |||
Db      4 PEF 6


RESULT   6
PT0709
T-cell receptor beta chain V-D-J region (161-2K) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0709
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0709
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-6 <FEE>
A;Experimental source: newborn thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             28.3%;  Score 15;  DB 2;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 DPE 6
          ||:
Db      4 DPD 6


RESULT   7
B24749
neuropeptide B - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 28-Jul-1987 #sequence_revision 28-Jul-1987 #text_change 31-Dec-1993
C;Accession: B24749
```

```
R;Yang, H.Y.T.; Fratta, W.; Majane, E.A.; Costa, E.
Proc. Natl. Acad. Sci. U.S.A. 82, 7757-7761, 1985
A;Title: Isolation, sequencing, synthesis, and pharmacological characterization of two b
A;Reference number: A94074; MUID:86067985
A;Accession: B24749
A;Molecule type: protein
A;Residues: 1-8 <YAN>
C;Keywords: neuropeptide


  Query Match             28.3%;  Score 15;  DB 2;  Length 8;
  Best Local Similarity   33.3%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLHDPE 6
             :|  |:
Db       1 FLFQPQ 6


RESULT   8
D58503
translation elongation factor EF-Tu - unidentified bacterium (fragment)
C;Species: unidentified bacterium
C;Date: 07-Feb-1997 #sequence_revision 07-Feb-1997 #text_change 28-May-1999
C;Accession: D58503
R;Binette, J.P.; Binette, M.B.
submitted to the Protein Sequence Database, October 1996
A;Description: The proteins of kidney and gallbladder stones.
A;Reference number: A58501
A;Accession: D58503
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <BIN>
A;Experimental source: human bile and stones
C;Superfamily: translation elongation factor Tu; translation elongation factor Tu homolo
C;Keywords: GTP binding


  Query Match             28.3%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 PEF 7
             |:|
Db       4 PQF 6


RESULT   9
A60108
exotoxin A - Streptococcus pyogenes (strain C203.S) (fragment)
N;Alternate names: blastogen A; scarlet fever toxin
C;Species: Streptococcus pyogenes
C;Date: 10-Nov-1992 #sequence_revision 10-Nov-1992 #text_change 07-Feb-1997
C;Accession: A60108
R;Schlievert, P.M.; Gray, E.D.
Infect. Immun. 57, 1865-1867, 1989
A;Title: Group A streptococcal pyrogenic exotoxin (scarlet fever toxin) type A and blast
A;Reference number: A60108; MUID:89254013
A;Accession: A60108
A;Molecule type: protein
A;Residues: 1-9 <SCH>
C;Keywords: exotoxin


  Query Match             28.3%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 DPE 6
             ||:
Db       2 DPD 4
```

```
RESULT  10
XGHUEU
urine glycopeptide - human
C;Species: Homo sapiens (man)
C;Date: 20-Jun-2000 #sequence_revision 20-Jun-2000 #text_change 20-Jun-2000
C;Accession: A03188
R;Lote, C.J.; Weiss, J.B.
Biochem. J. 123, 25P, 1971
A;Title: Identification in urine of a low-molecular-weight polar glycopeptide contain
A;Reference number: A03188; MUID:72062338
A;Accession: A03188
A;Molecule type: protein
A;Residues: 1-8 <LOT>
C;Comment: The identity of the glycoprotein from which this peptide is derived is unk
re has also been found (see PIR:XGHUE).
C;Superfamily: unassigned animal peptides
C;Keywords: glycoprotein
F;1/Binding site: carbohydrate (Cys) (covalent) #status experimental


  Query Match             26.4%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 HD 4
             ||
Db       5 HD 6


RESULT  11
A28719
thymic humoral factor gamma-2 - bovine (fragment)
C;Species: Bos primigenius taurus (cattle)
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 18-Jun-1993
C;Accession: A28719
R;Burstein, Y.; Buchner, V.; Pecht, M.; Trainin, N.
Biochemistry 27, 4066-4071, 1988
A;Title: Thymic humoral factor gamma-2: purification and amino acid sequence of an im
A;Reference number: A28719; MUID:88326920
A;Accession: A28719
A;Molecule type: protein
A;Residues: 1-8 <BUR>


  Query Match             26.4%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 PEF 7
             |:|
Db       5 PKF 7


RESULT  12
PT0231
Ig heavy chain CDR3 region (clone 1-118B) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C;Accession: PT0231
R;Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991
A;Title: Preferential utilization of specific immunoglobulin heavy chain diversity an
A;Reference number: PT0222; MUID:91108337
A;Accession: PT0231
A;Molecule type: DNA
A;Residues: 1-9 <YAM>
A;Experimental source: B lymphocyte
C;Keywords: heterotetramer; immunoglobulin


  Query Match             26.4%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy       1 YLH 3
           | |
Db       1 YTH 3


RESULT  13
I51049
metallothionein-A - rainbow trout (fragment)
C;Species: Oncorhynchus mykiss (rainbow trout)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Sep-1996
C;Accession: I51049
R;Olsson, P.E.; Kling, P.; Erkell, L.J.; Kille, P.
Eur. J. Biochem. 230, 344-349, 1995
A;Title: Structural and functional analysis of the rainbow trout (Oncorhyncus mykiss) me
A;Reference number: I51049; MUID:95324545
A;Accession: I51049
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4 <OLS>
A;Cross-references: EMBL:X80181; NID:g1019799


  Query Match             24.5%;  Score 13;  DB 2;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 DP 5
           ||
Db       2 DP 3


RESULT  14
B60274
major protein antigen MPT32 - Mycobacterium tuberculosis (fragment)
C;Species: Mycobacterium tuberculosis
C;Date: 11-Dec-1992 #sequence_revision 11-Dec-1992 #text_change 30-Sep-1993
C;Accession: B60274
R;Nagai, S.; Wiker, H.G.; Harboe, M.; Kinomoto, M.
Infect. Immun. 59, 372-382, 1991
A;Title: Isolation and partial characterization of major protein antigens in the culture
A;Reference number: A60274; MUID:91099989
A;Accession: B60274
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-5 <NAG>


  Query Match             24.5%;  Score 13;  DB 2;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 DP 5
           ||
Db       1 DP 2


RESULT  15
PT0665
T-cell receptor beta chain V-D-J region (121-3BM) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0665
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0665
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-7 <FEE>
A;Experimental source: day 4 postnatal thymus, strain BALB/c
C;Keywords: T-cell receptor
```

```
  Query Match             24.5%;  Score 13;  DB 2;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 DPEF 7
           | :|
Db       4 DADF 7


RESULT  16
S71919
alcohol dehydrogenase (EC 1.1.1.1) - grass carp (fragment)
C;Species: Ctenopharyngodon idella (grass carp)
C;Date: 14-Apr-1998 #sequence_revision 24-Apr-1998 #text_change 07-May-1999
C;Accession: S71919
R;Tsui, H.T.; Mock, W.Y.; Lau, K.K.; Fong, W.P.
Biochim. Biophys. Acta 1296, 41-46, 1996
A;Title: Proteolytic activation of grass carp (Ctenopharygodon idellus) liver alcohol
A;Reference number: S71919; MUID:96350418
A;Accession: S71919
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-8 <TSU>
A;Note: the source is designated Ctenopharyngodon idellus
C;Keywords: NAD; oxidoreductase


  Query Match             24.5%;  Score 13;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 DP 5
           ||
Db       2 DP 3


RESULT  17
PT0368
Ig gamma chain C region (gamma-1) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Aug-1996
C;Accession: PT0368
R;Milili, M.; Fougereau, M.; Guglielmi, P.; Schiff, C.
Mol. Immunol. 28, 753-761, 1991
A;Title: Early occurrence of immunoglobulin isotype switching in human fetal liver.
A;Reference number: PT0368; MUID:91312348
A;Accession: PT0368
A;Molecule type: mRNA
A;Residues: 1-8 <MIL>
A;Experimental source: fetal liver
C;Keywords: immunoglobulin


  Query Match             24.5%;  Score 13;  DB 2;  Length 8;
  Best Local Similarity   37.5%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 LHDPEFNL 9
           ||     |
Db       1 LHQGPIGL 8


RESULT  18
B45796
dihydrolipoamide S-succinyltransferase (EC 2.3.1.61) - Pseudomonas fluorescens (fragm
C;Species: Pseudomonas fluorescens
C;Date: 14-May-1993 #sequence_revision 14-May-1993 #text_change 05-May-2000
C;Accession: B45796
R;Benen, J.A.E.; Van Berkel, W.J.H.; Van Dongen, W.M.A.M.; Mueller, F.; De Kok, A.
J. Gen. Microbiol. 135, 1787-1797, 1989
A;Title: Molecular cloning and sequence determination of the lpd gene encoding lipoam
```

```
A;Reference number: A45796; MUID:90132584
A;Accession: B45796
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-9 <BEN>
A;Cross-references: GB:M28356; NID:g151343; PIDN:AAA99233.1; PID:g151344
C;Superfamily: dihydrolipoamide acetyltransferase; lipoyl/biotin-binding homology
C;Keywords: acyltransferase; coenzyme A; lipoamide; tricarboxylic acid cycle


  Query Match             24.5%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 DP 5
           ||
Db       1 DP 2


RESULT  19
G58502
kidney and bladder stone protein - unidentified bacterium (fragment)
C;Species: unidentified bacterium
C;Date: 07-Feb-1997 #sequence_revision 07-Feb-1997 #text_change 10-Jul-1998
C;Accession: G58502
R;Binette, J.P.; Binette, M.B.
submitted to the Protein Sequence Database, October 1996
A;Description: The proteins of kidney and gallbladder stones.
A;Reference number: A58501
A;Accession: G58502
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <BIN>
A;Experimental source: human kidney stone, bladder stone
A;Note: a secondary sequence AAKENPXD was also found


  Query Match             24.5%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LHDPEF 7
           | | :|
Db       2 LPDVKF 7


RESULT  20
S36898
ribosomal protein S8 - Mycobacterium bovis (fragment)
C;Species: Mycobacterium bovis
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Jan-1995
C;Accession: S36898
R;Ohara, N.; Kimura, M.; Higashi, Y.; Yamada, T.
FEBS Lett. 331, 9-14, 1993
A;Title: Isolation and amino acid sequence of the 30S ribosomal protein S19 from Mycobac
A;Reference number: S36887; MUID:94009653
A;Accession: S36898
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <OHA>
C;Keywords: protein biosynthesis; ribosome


  Query Match             24.5%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 DP 5
           ||
Db       4 DP 5


RESULT  21
PT0080
60K Ca binding protein - edible frog (fragment)
C;Species: Rana esculenta (edible frog)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C;Accession: PT0080
R;Treveso, S.; Zorzato, F.; Chiozzi, P.; Melandri, P.; Volpe, P.; Pozzan, T.
Biochem. Biophys. Res. Commun. 175, 444-450, 1991
A;Title: Frog brain expresses a 60 kDa Ca2+ binding protein similar to mammalian calr
A;Reference number: PT0080; MUID:91207333
A;Accession: PT0080
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <TRE>


  Query Match             24.5%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 DP 5
           ||
Db       1 DP 2


RESULT  22
G56978
collagen alpha 1(II) chain - bovine (fragment)
N;Alternate names: collagen alpha 3(XI) chain
C;Species: Bos primigenius taurus (cattle)
C;Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 03-Oct-1995
C;Accession: G56978
R;Wu, J.J.; Eyre, D.R.
J. Biol. Chem. 270, 18865-18870, 1995
A;Title: Structural analysis of cross-linking domains in cartilage type XI collagen.
A;Reference number: A56978; MUID:95370194
A;Accession: G56978
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <WUA>
A;Note: the residue designated 'X' is modified lysine in collagen 3(XI) some cross-li


  Query Match             24.5%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 DP 5
           ||
Db       8 DP 9


RESULT  23
PD0443
3-oxoacid CoA-transferase (EC 2.8.3.5) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 05-Feb-1999
C;Accession: PD0443
R;Kawakami, T.; Uchida, T.; Sakai, T.; Kamo, M.; Morimasa, T.; Tsugita, A.
submitted to JIPID, August 1998
A;Description: Proteome analysis of mouse brain.
A;Reference number: PD0441
A;Contents: Striatum
A;Accession: PD0443
A;Molecule type: protein
A;Residues: 1-9 <KAW>
C;Keywords: CoA-transferase


  Query Match             24.5%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 DP 5
```

```
          ||
Db       6 DP 7


RESULT  24
S39437
D-amino-acid oxidase (EC 1.4.3.3) - Trigonopsis variabilis (fragment)
C;Species: Trigonopsis variabilis
C;Date: 19-Mar-1997 #sequence_revision 05-Dec-1997 #text_change 07-May-1999
C;Accession: S39437
R;Schraeder, T.; Andreesen, J.R.
Eur. J. Biochem. 218, 735-744, 1993
A;Title: Evidence for the functional importance of Cys298 in D-amino acid oxidase from T
A;Reference number: S39437; MUID:94094869
A;Accession: S39437
A;Molecule type: protein
A;Residues: 1-9 <SCH>
A;Experimental source: CBS 4095
C;Function:
A;Description: oxidoreductase; catalyzes the oxidation of D-amino acids to their corresp
A;Note: reoxidation of the enzyme by molecular oxygen is accompanied by the release of h
C;Keywords: FAD; oxidoreductase


  Query Match             24.5%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 HDP 5
           | |
Db       4 HRP 6


RESULT  25
S53508
starvation-induced ribonuclease - tomato
C;Species: Lycopersicon esculentum (tomato)
C;Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 07-May-1999
C;Accession: S53508
R;Koeck, M.; Loeffler, A.; Abel, S.; Glund, K.
Plant Mol. Biol. 27, 477-485, 1995
A;Title: cDNA structure and regulatory properties of a family of starvation-induced rib
A;Reference number: S53506; MUID:95201242
A;Accession: S53508
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-4 <KOE>


  Query Match             22.6%;  Score 12;  DB 2;  Length 4;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 PEF 7
           | |
Db       2 PTF 4


RESULT  26
I54357
schwannomin - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C;Accession: I54357
R;Huynh, D.P.; Nechiporuk, T.; Pulst, S.
Hum. Mol. Genet. 3, 1075-1079, 1994
A;Title: Alternative transcripts in the mouse neurofibromatosis type 2 (NF2) gene are c
A;Reference number: I54357; MUID:95072570
A;Accession: I54357
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-4 <RES>
A;Cross-references: GB:L28838; NID:g454836; PIDN:AAA57150.1; PID:g601923
```

```
C;Genetics:
A;Gene: NF2

  Query Match             22.6%;  Score 12;  DB 2;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 PE 6
           ||
Db       2 PE 3


RESULT  27
A44692
fulicin - giant African snail
C;Species: Achatina fulica (giant African snail)
C;Date: 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 11-Jul-1997
C;Accession: A44692
R;Ohta, N.; Kubota, I.; Takao, T.; Shimonishi, Y.; Yasuda-Kamatani, Y.; Minakata, H.;
Biochem. Biophys. Res. Commun. 178, 486-493, 1991
A;Title: Fulicin, a novel neuropeptide containing a D-amino acid residue isolated fro
A;Reference number: A44692; MUID:91315471
A;Accession: A44692
A;Molecule type: protein
A;Residues: 1-5 <OHT>
C;Keywords: amidated carboxyl end; D-amino acid; neuropeptide
F;2/Modified site: D-asparagine (Asn) #status experimental
F;5/Modified site: amidated carboxyl end (Val) #status experimental


  Query Match             22.6%;  Score 12;  DB 2;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 FN 8
           ||
Db       1 FN 2


RESULT  28
JN0861
peptidyl-dipeptidase A inhibitory peptide C111 - striped bonito
C;Species: Sarda orientalis (striped bonito)
C;Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 07-May-1999
C;Accession: JN0861
R;Matsumura, N.; Fujii, M.; Takeda, Y.; Shimizu, T.
Biosci. Biotechnol. Biochem. 57, 1743-1744, 1993
A;Title: Isolation and characterization of angiotensin I-converting enzyme inhibitory
A;Reference number: JN0859; MUID:94080036
A;Accession: JN0861
A;Molecule type: protein
A;Residues: 1-6 <MAT>
A;Experimental source: liver
C;Comment: The carboxyl end is essential for the protein's expression of angiotensin
C;Superfamily: bradykinin-potentiating peptide
C;Keywords: angiotensin-converting enzyme inhibitor


  Query Match             22.6%;  Score 12;  DB 2;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLH 3
           | |
Db       3 YPH 5


RESULT  29
S11556
hydrogensulfite reductase (EC 1.8.99.3) chain 2 - Desulfovibrio thermophilus (fragmen
N;Alternate names: bisulfite reductase; desulfofuscidin
C;Species: Desulfovibrio thermophilus
```

```
C;Date: 19-Mar-1997 #sequence_revision 12-Dec-1997 #text_change 30-Jan-1998
C;Accession: S11556
R;Fauque, G.; Lino, A.R.; Czechowski, M.; Kang, L.; DerVartanian, D.V.; Moura, J.J.G.;
Biochim. Biophys. Acta 1040, 112-118, 1990
A;Title: Purification and characterization of bisulfite reductase (desulfofuscidin) fro
A;Reference number: S11024; MUID:90335276
A;Accession: S11556
A;Molecule type: protein
A;Residues: 1-6 <FAU>
C;Keywords: oxidoreductase


  Query Match             22.6%;  Score 12;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 PE 6
            ||
Db        1 PE 2


RESULT  30
S78764
ribosomal protein MRP-S23, mitochondrial - bovine (fragment)
C;Species: Bos primigenius taurus (cattle)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: S78764
R;Graack, H.R.
submitted to the Protein Sequence Database, July 1999
A;Reference number: S78760
A;Accession: S78764
A;Molecule type: protein
A;Residues: 1-6 <GRA>
C;Keywords: mitochondrion
F;1-6/Product: ribosomal protein MRP-S23 (fragment) #status experimental <MAT>


  Query Match             22.6%;  Score 12;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LH 3
            ||
Db        1 LH 2


RESULT  31
PQ0663
membrane protein - porcine epidemic diarrhea virus (isolate Belgian CV777) (fragment)
C;Species: porcine epidemic diarrhea virus
C;Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 08-Oct-1999
C;Accession: PQ0663
R;Bridgen, A.; Duarte, M.; Tobler, K.; Laude, H.; Ackermann, M.
J. Gen. Virol. 74, 1795-1804, 1993
A;Title: Sequence determination of the nucleocapsid protein gene of the porcine epidemi
issible gastroenteritis virus.
A;Reference number: JQ2191; MUID:93389433
A;Accession: PQ0663
A;Molecule type: mRNA
A;Residues: 1-7 <BRI>
A;Cross-references: GB:Z14976; NID:g311650; PIDN:CAA78699.1; PID:g584083
C;Comment: This virus is coronavirus related to human coronavirus 229E.
C;Keywords: membrane protein


  Query Match             22.6%;  Score 12;  DB 2;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LH 3
            ||
Db        4 LH 5
```

```
RESULT  32
H33098
180K exoantigen - malaria parasite (Plasmodium falciparum) (fragments)
C;Species: Plasmodium falciparum
C;Date: 24-Aug-1990 #sequence_revision 24-Aug-1990 #text_change 09-Jun-2000
C;Accession: H33098
R;Nichols, J.H.; Hager, L.P.
submitted to the Protein Sequence Database, May 1990
A;Reference number: A33098
A;Accession: H33098
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-7 <NIC>


  Query Match             22.6%;  Score 12;  DB 2;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        7 FN 8
            ||
Db        3 FN 4


RESULT  33
E48394
glycoprotein component 16/major fat-globule membrane protein/MFG-E8 homolog - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 07-Feb-1997
C;Accession: E48394
R;Mather, I.H.; Banghart, L.R.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A;Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-
II-like sequences.
A;Reference number: A48394; MUID:93250576
A;Accession: E48394
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-7 <MAT>
A;Experimental source: milk
A;Note: sequence extracted from NCBI backbone (NCBIP:131450)
C;Keywords: glycoprotein


  Query Match             22.6%;  Score 12;  DB 2;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 PE 6
            ||
Db        3 PE 4


RESULT  34
B39040
calsequestrin, fast skeletal muscle - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 23-Aug-1991 #sequence_revision 23-Aug-1991 #text_change 23-Feb-1997
C;Accession: B39040
R;Cala, S.E.; Jones, L.R.
J. Biol. Chem. 266, 391-398, 1991
A;Title: Phosphorylation of cardiac and skeletal muscle calsequestrin isoforms by cas
A;Reference number: A39040; MUID:91093153
A;Accession: B39040
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-7 <CAL>
C;Keywords: phosphoprotein; skeletal muscle


  Query Match             22.6%;  Score 12;  DB 2;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
```

```
  Matches     3;  Conservative     1;  Mismatches     2;  Indels     0;  Gaps     0;

Qy       4 DPEFNL 9
           | | :|
Db       1 DDEEDL 6


RESULT  35
I48293
transforming protein cKi-ras - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C;Accession: I48293
R;Trusko, S.P.; Hoffman, E.K.; George, D.L.
Nucleic Acids Res. 17, 9259-9265, 1989
A;Title: Transcriptional activation of cKi-ras proto-oncogene resulting from retroviral
A;Reference number: I48293; MUID:90067925
A;Accession: I48293
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-7 <RES>
A;Cross-references: EMBL:X15888; NID:g50415; PIDN:CAA33898.1; PID:g581942

  Query Match             22.6%;  Score 12;  DB 2;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches     2;  Conservative     1;  Mismatches     1;  Indels     0;  Gaps     0;

Qy       6 EFNL 9
           |: |
Db       3 EYKL 6


Search completed: December 16, 2000, 03:35:13
Job time: 5646 sec
```

```
                         GenCore version 4.5
                Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 03:07:20 ; Search time 107.12 Seconds
                                              (without alignments)
                                               2.873 Million cell updates/sec

Title:           US-09-529-121-10
Perfect score:   53
Sequence:        1 YLHDPEFNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters:      61695

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       A_Geneseq_36:*
                1:  /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
                2:  /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
                3:  /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
                4:  /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
                5:  /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
                6:  /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
                7:  /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
                8:  /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
                9:  /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
               10:  /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
               11:  /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
               12:  /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
               13:  /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
               14:  /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
               15:  /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
               16:  /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
               17:  /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
               18:  /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
               19:  /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
               20:  /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
               21:  /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                               SUMMARIES
                 %
Result          Query
  No.   Score   Match  Length DB  ID                       Description
--------------------------------------------------------------------------
    1      25    47.2       6  20  W84423                  HIV-1 nucleic acid
    2      25    47.2       7  20  Y16929                  Heat shock protein
    3      25    47.2       8  18  W38402                  Synthetic pMEL17 p
    4      25    47.2       9  16  R82119                  Melanoma-specific
    5      23    43.4       6  19  W87278                  Peptide determined
    6      23    43.4       6  19  W87245                  Peptide determined
    7      23    43.4       8  19  W61745                  Peptide mimetic of
    8      23    43.4       9  20  Y47777                  Immunogenic peptid
    9      23    43.4       9  21  Y77439                  Fibronectin CS1-de
   10      23    43.4       9  21  Y77440                  Fibronectin CS1-de
   11      22    41.5       4  19  W52752                  Aminomethylcoumari
   12      22    41.5       5  12  R14525                  Interleukin-1 beta
   13      22    41.5       5  14  R34850                  Interleukin 1-beta
   14      22    41.5       5  20  Y21732                  Caspase sensitive
   15      22    41.5       6  10  P93063                  Peptide encoded by
   16      22    41.5       8  14  R34855                  IL-1beta protease
   17      21    39.6       5  12  R12720                  Pentapeptide paral
   18      21    39.6       5  15  R51584                  Mimotope peptide #
   19      21    39.6       5  16  R69952                  Pentameric mimotop
   20      21    39.6       5  17  R98698                  Peptide 78 from 88
   21      21    39.6       5  20  Y23613                  SEQ ID 104 of US59
   22      21    39.6       7  13  R20337                  Sequence of synthe
   23      21    39.6       7  13  R20345                  Sequence of synthe
   24      21    39.6       7  15  R55584                  AmEPV spheroidin p
   25      21    39.6       7  20  Y17008                  Heat shock protein
   26      21    39.6       8  20  Y16837                  Heat shock protein
   27      21    39.6       9  15  R67767                  N-terminal sequenc
   28      21    39.6       9  20  Y48130                  Immunogenic peptid
   29      21    39.6       9  20  Y23570                  N-terminal sequene
   30      21    39.6       9  20  W89129                  ScFV (dig) variant
   31      20    37.7       6  19  W86288                  Rodent interleukin
   32      20    37.7       6  19  W87325                  Peptide determined
   33      20    37.7       6  19  W87276                  Peptide determined
   34      20    37.7       7  16  R76987                  Immune stimulating
   35      20    37.7       7  17  W07489                  Soybean pepsin dig
   36      20    37.7       7  19  W52610                  Disulphide-circula
   37      20    37.7       7  20  Y23850                  Peptide derived fr
   38      20    37.7       7  21  W90982                  N. denitrificans a
   39      20    37.7       8  16  R73504                  Human TSH receptor
   40      20    37.7       8  17  W00487                  Peptide epitope of
   41      20    37.7       8  17  R86587                  Autotaxin peptide
   42      20    37.7       8  20  Y17077                  Human platelet hep
   43      20    37.7       9  15  R52611                  Inhibitor of signa
   44      20    37.7       9  16  R72567                  Penicillium purpur
   45      20    37.7       9  18  W15781                  Protein kinase C-b
   46      20    37.7       9  19  W54879                  Isozyme-specific a
   47      20    37.7       9  20  Y46637                  Immunogenic peptid
   48      20    37.7       9  20  Y48166                  Immunogenic peptid
   49      20    37.7       9  20  Y09526                  Carcinoembryonic a
   50      20    37.7       9  20  Y09527                  Carcinoembryonic a
   51      20    37.7       9  20  W92364                  H. medicinalis FXa
   52      20    37.7       9  21  Y85445                  Tumour antigen pro
   53      19    35.8       5  14  R37032                  Pentapeptide compo
   54      19    35.8       6  14  R37394                  Peptide for treati
   55      19    35.8       6  14  R42754                  Consensus peptide,
   56      19    35.8       6  17  R96771                  Synthetic library
   57      19    35.8       6  17  R96774                  Synthetic library
   58      19    35.8       6  19  W87125                  Peptide determined
   59      19    35.8       6  19  W87158                  Peptide determined
   60      19    35.8       6  21  Y83963                  Human interleukin
   61      19    35.8       7  12  R15328                  Amino acid copper
   62      19    35.8       7  12  R15330                  Amino acid copper
   63      19    35.8       7  12  R15332                  Amino acid copper
   64      19    35.8       7  12  R15336                  Amino acid copper
   65      19    35.8       7  15  R55100                  Prostate-specific
   66      19    35.8       7  17  W02239                  Prostate-specific
   67      19    35.8       7  19  W65638                  Angiotensin III an
   68      19    35.8       7  20  Y48917                  Membrane dipeptida
   69      19    35.8       7  20  Y18098                  Sel-1L protein PTB
   70      19    35.8       7  20  Y01258                  US5851813 peptide
   71      19    35.8       7  21  W90810                  Rat intestinal muc
   72      19    35.8       8  12  R11375                  Polypeptide for tr
   73      19    35.8       8  12  R11380                  Polypeptide for tr
   74      19    35.8       8  15  Y38229                  HIV-derived HLA-bi
   75      19    35.8       8  16  R73505                  Human TSH receptor


                              ALIGNMENTS


RESULT   1
W84423
ID   W84423 standard; Peptide; 6 AA.
XX
AC   W84423;
```

```
XX
DT   22-MAR-1999  (first entry)
XX
DE   HIV-1 nucleic acid binding protein zinc finger 2 peptide.
XX
KW   zinc finger; nucleotide-binding protein; cell proliferative disorder;
KW   gene therapy; cancer; psoriasis; pemphigus vulgaris; Bechet's syndrome;
KW   lipid histiocytosis; human immune deficiency virus; HIV; viral infection;
KW   transgenic plant.
XX
OS   Synthetic.
XX
PN   WO9854311-A1.
XX
PD   03-DEC-1998.
XX
PF   27-MAY-1998;   98WO-US10801.
XX
PR   27-MAY-1997;   97US-0863813.
XX
PA   (SCRI ) SCRIPPS RES INST.
XX
PI   Barbas CF,  Gottesfeld JM,  Wright PE;
XX
DR   WPI; 1999-059831/05.
XX
PT   New zinc finger nucleotide-binding protein variant that modulates
PT   selected nucleotide sequence - used for treatment of proliferative
PT   and viral diseases by gene therapy, and can be made selective for
PT   any target sequence
XX
PS   Example 10; Fig 9; 158pp; English.
XX
CC   W84422-41 represent the peptides obtained from randomised finger 2
CC   sequences of zinc finger nucleotide-binding proteins that bind to HIV-1
CC   target sequences. The peptides were produced in the course of the
CC   invention. The specification describes zinc finger nucleotide-binding
CC   protein variants with at least two zinc finger modules that bind to a
CC   cellular nucleotide sequence and modulate its function. Zinc finger
CC   proteins, and compositions containing them, are used to increase or
CC   reduce transcription of a gene linked to the cellular nucleotide
CC   sequence. The proteins are used specifically for treating or preventing
CC   cell proliferative disorders (in humans, animals or plants, including
CC   those induced by viruses), particularly where expressed from nucleic acid
CC   by gene therapy (including ex vivo methods). Typical diseases that can be
CC   treated are many forms of cancer, psoriasis, pemphigus vulgaris,
CC   Bechet's syndrome and lipid histiocytosis, also treatment of human
CC   immune deficiency virus (HIV) and other viral infections, and production
CC   of transgenic plants resistant to bacterial and viral diseases. The
CC   present oligonucleotide is used in the course of the invention.
XX
SQ   Sequence   6 AA;

  Query Match             47.2%;  Score 25;  DB 20;  Length 6;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLHDP 5
            ||| |
Db        2 ylhtp 6

RESULT   2
Y16929
ID   Y16929 standard; peptide; 7 AA.
XX
AC   Y16929;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Heat shock protein (hsp) binding peptide.
XX
KW   Conjugate peptide; heat shock protein; hsp; phage display library; virus;
KW   surface protein; tethering peptide; chaperone process; cytokine; cancer;
KW   neoplastic disease; infectious disease; bacterium; immune system; fungus;
KW   acquired immune deficiency; autoimmune disease.
XX
OS   Synthetic.
XX
PN   WO9922761-A1.
XX
PD   14-MAY-1999.
XX
PF   22-OCT-1998;   98WO-US22335.
XX
PR   31-OCT-1997;   97US-0961707.
XX
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI   Hartl U,  Hoe MH,  Houghton A,  Mayhew M,  Moroi Y;
PI   Ouerfelli O,  Rothman JE;
XX
DR   WPI; 1999-313177/26.
XX
PT   Identifying peptides which bind heat shock proteins
XX
PS   Examples; Page 21; 155pp; English.
XX
CC   The invention relates to conjugate peptides engineered to noncovalently
CC   bind to heat shock proteins (hsp). A method of identifying a hsp binding
CC   peptide comprises (a) contacting a phage display library having
CC   bacteriophage expressing, in a surface protein, inserted peptides with a
CC   hsp target, and bound to a benzoquinone ansamycin antibiotic (BAA), in a
CC   physiologic binding buffer; (b) isolating a phage binding to the hsp
CC   target; and (c) identifying the inserted peptide expressed. The peptides
CC   which bind to a hsp can be used as tethering peptides for a hsp which may
CC   serve as an accessory in a chaperone process and/or may comprise a
CC   cytokine. They can also be coupled to antigens to induce an immune
CC   response. Such compositions can be used for treating neoplastic disease,
CC   e.g. cancers, infectious diseases, e.g. diseases caused by a bacterium,
CC   virus, protozoan, mycoplasma, fungus, yeast, parasite or prion, or a
CC   disease of the immune system, e.g. acquired immune deficiencies or
CC   autoimmune diseases.
XX
SQ   Sequence   7 AA;

  Query Match             47.2%;  Score 25;  DB 20;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LHDP 5
            ||||
Db        3 lhdp 6

RESULT   3
W38402
ID   W38402 standard; peptide; 8 AA.
XX
AC   W38402;
XX
DT   08-APR-1998  (first entry)
XX
DE   Synthetic pMEL17 peptide.
XX
KW   Melanoma; immunogen; cytotoxic T lymphocyte; CTL;
KW   human leukocyte antigen-A1; HLA-A1; human leukocyte antigen-A3;
KW   HLA-A3; epitope; pMEL-17; tyrosinase; vaccine; protection.
XX
OS   Synthetic.
XX
PN   WO9734613-A1.
XX
PD   25-SEP-1997.
```

```
XX
PF   17-MAR-1997;   97WO-US04958.
XX
PR   04-OCT-1996;   96US-0027627.
PR   19-MAR-1996;   96US-0013972.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI   Cox AL,  Engelhard VH,  Hendrikson RC,  Hunt DF,  Kittlesen D;
PI   Shabanowitz J,  Skipper J,  Slingluff CL;
XX
DR   WPI; 1997-479982/44.
XX
PT   Melanoma-specific immunogens of pMel-17 and tyrosinase - useful in
PT   vaccination for producing melanoma-specific cytotoxic T lymphocytes
XX
PS   Example 9; Page 65; 106pp; English.
XX
CC   The present peptide was used in the preparation of a novel melanoma
CC   specific immunogen, comprising at least 1 melanoma specific
CC   cytotoxic T lymphocyte (CTL) epitope, where at least 1 of the
CC   epitopes is substantially homologous to a human leukocyte
CC   antigen-A1 (HLA-A1) and HLA-A3 restricted epitope of a melanoma
CC   antigen, either pMEL-17 or tyrosinase. The immunogen can be used in
CC   vaccines for protection against melanoma in mammals.
XX
SQ   Sequence   8 AA;


  Query Match              47.2%;  Score 25;  DB 18;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LHDP 5
            ||||
Db        2 lhdp 5


RESULT   4
R82119
ID   R82119 standard; peptide; 9 AA.
XX
AC   R82119;
XX
DT   25-MAR-1996  (first entry)
XX
DE   Melanoma-specific mutant immunogen epitope 9mer peptide.
XX
KW   Melanoma; immunogen; epitope; homologue; vaccine; immunotherapy;
KW   cytotoxic T cell; lymphocyte; HLA-A2.
XX
OS   Homo sapiens.
XX
PN   WO9522561-A2.
XX
PD   24-AUG-1995.
XX
PF   16-FEB-1995;   95WO-US01991.
XX
PR   29-APR-1994;   94US-0234784.
PR   16-FEB-1994;   94US-0197399.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI   Cox AL,  Engelhard VH,  Hunt DF,  Shabanowitz J,  Slingluff CL;
XX
DR   WPI; 1995-302688/39.
XX
PT   Melanoma-specific immunogen comprises epitope(s) homologous with
PT   pMel.17 - are highly potent stimulators of HLA-A2+CTL's useful in
PT   adoptive immuno-therapy
XX
PS   Example 8; Page 51; 148pp; English.
```

```
XX
CC   A melanoma-specific immunogen homologous with pMel-17 comprises one
CC   or more CTL (cytotoxic T lymphocyte) epitopes from the group R82098-
CC   R82194 capable of eliciting a CTL response. The epitopes R82098-
CC   R82108 are of particular interest. The immunogen can be used for
CC   partial protection in mammals against melanoma peptides which are
CC   homologous with pMel-17 are highly potent stimulators of HLA-A2+
CC   CTLs in several cell lines and can be used in immunotherapy or
CC   incorporated into immunogenic conjugates as vaccines.
XX
SQ   Sequence   9 AA;


  Query Match              47.2%;  Score 25;  DB 16;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LHDP 5
            ||||
Db        2 lhdp 5


RESULT   5
W87278
ID   W87278 standard; peptide; 6 AA.
XX
AC   W87278;
XX
DT   09-FEB-1999  (first entry)
XX
DE   Peptide determined by the method of the invention.
XX
KW   Amino acid determination; molecular mass; fragmentation spectrum;
KW   DNA cloning; anti-body; recombinant; modification; mass spectrometry.
XX
OS   Synthetic.
XX
PN   GB2325465-A.
XX
PD   25-NOV-1998.
XX
PF   22-MAY-1998;   98GB-0011196.
XX
PR   22-MAY-1997;   97GB-0010582.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Parekh RB,  Prime SB,  Townsend RR,  Wedd NS;
XX
DR   WPI; 1998-571195/49.
XX
PT   Peptide sequence determination used in e.g. DNA cloning - by
PT   comparing mass spectra of the unknown peptide with a library of
PT   linear chain known peptide sequences
XX
PS   Example 1; Page 20; 40pp; English.
XX
CC   The invention relates to a method for determination of the amino acid
CC   sequence of an unknown peptide. The method comprises (a) determining
CC   the molecular mass and an experimental fragmentation spectrum for the
CC   peptide; (b) comparing the experimental fragmentation spectrum of the
CC   unknown peptide with a theoretical fragmentation spectra calculated for
CC   a peptide library composed of all possible linear sequences of amino
CC   acids having a total mass that corresponds to the molecular mass of the
CC   unknown peptide; and (c) identifying a peptide in the library with a
CC   theoretical fragmentation spectrum that most closely matches the
CC   fragmentation spectrum of the unknown peptide. The method is useful in
CC   DNA cloning, anti-body production, identification of recombinant
CC   products, and the study of post-translational modifications. It allows
CC   the sequence of unknown peptides or proteins with no sub-sequence
CC   identity, to be characterised using mass spectrometry. Sequences W87101
CC   to W87364 represent a library of linear peptides constructed to exemplify
CC   the method. The isoleucine residue in these peptides can be replaced by
```

```
CC   leucine to construct another 264 linear peptides to be included in the
CC   library.
XX
SQ   Sequence   6 AA;


  Query Match             43.4%;  Score 23;  DB 19;  Length 6;
  Best Local Similarity   75.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLHD 4
           |:||
Db       2 yihd 5


RESULT   6
W87245
ID   W87245 standard; peptide; 6 AA.
XX
AC   W87245;
XX
DT   09-FEB-1999  (first entry)
XX
DE   Peptide determined by the method of the invention.
XX
KW   Amino acid determination; molecular mass; fragmentation spectrum;
KW   DNA cloning; anti-body; recombinant; modification; mass spectrometry.
XX
OS   Synthetic.
XX
PN   GB2325465-A.
XX
PD   25-NOV-1998.
XX
PF   22-MAY-1998;   98GB-0011196.
XX
PR   22-MAY-1997;   97GB-0010582.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Parekh RB,  Prime SB,  Townsend RR,  Wedd NS;
XX
DR   WPI; 1998-571195/49.
XX
PT   Peptide sequence determination used in e.g. DNA cloning - by
PT   comparing mass spectra of the unknown peptide with a library of
PT   linear chain known peptide sequences
XX
PS   Example 1; Page 20; 40pp; English.
XX
CC   The invention relates to a method for determination of the amino acid
CC   sequence of an unknown peptide. The method comprises (a) determining
CC   the molecular mass and an experimental fragmentation spectrum for the
CC   peptide; (b) comparing the experimental fragmentation spectrum of the
CC   unknown peptide with a theoretical fragmentation spectra calculated for
CC   a peptide library composed of all possible linear sequences of amino
CC   acids having a total mass that corresponds to the molecular mass of the
CC   unknown peptide; and (c) identifying a peptide in the library with a
CC   theoretical fragmentation spectrum that most closely matches the
CC   fragmentation spectrum of the unknown peptide. The method is useful in
CC   DNA cloning, anti-body production, identification of recombinant
CC   products, and the study of post-translational modifications. It allows
CC   the sequence of unknown peptides or proteins with no sub-sequence
CC   identity, to be characterised using mass spectrometry. Sequences W87101
CC   to W87364 represent a library of linear peptides constructed to exemplify
CC   the method. The isoleucine residue in these peptides can be replaced by
CC   leucine to construct another 264 linear peptides to be included in the
CC   library.
XX
SQ   Sequence   6 AA;


  Query Match             43.4%;  Score 23;  DB 19;  Length 6;
```

```
  Best Local Similarity   75.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLHD 4
           |:||
Db       1 yihd 4


RESULT   7
W61745
ID   W61745 standard; peptide; 8 AA.
XX
AC   W61745;
XX
DT   09-NOV-1998  (first entry)
XX
DE   Peptide mimetic of cytokine receptor gamma chain 12.
XX
KW   Peptide mimetic; gamma-chain; cytokine receptor; signal transduction;
KW   autoimmune disease; graft vs. host disease; transplant rejection;
KW   graft rejection; interleukin; immunosuppressant; T cell; B cell.
XX
OS   Synthetic.
XX
PN   WO9834631-A1.
XX
PD   13-AUG-1998.
XX
PF   06-FEB-1998;   98WO-US02339.
XX
PR   07-FEB-1997;   97US-0036941.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Korngold R,  Townsend RM;
XX
DR   WPI; 1998-446944/38.
XX
PT   New peptide(s) mimicking a loop in the gamma chain of cytokine
PT   receptors - inhibit signal transduction through these receptors,
PT   useful as immunosuppressants for treating or preventing e.g.
PT   leukaemia, autoimmune disease, graft rejection
XX
PS   Claim 2; Page 28; 54pp; English.
XX
CC   The peptide mimetics W61734-W61756 and W77001-W77014 mimic a loop on the
CC   gamma-chain of cytokine receptors, and interacts with a cytokine or a
CC   gamma-chain partner receptor chain of a heterodimeric cytokine receptor.
CC   They inhibit signal transduction mediated by cytokine:receptor binding
CC   (of cytokines that bind to receptors with a gamma-chain).  They are used
CC   to inhibit or suppress cytokine-mediated immune responses, growth,
CC   proliferation, function and activity of cells. Particularly they are used
CC   for treatment or prevention of lymphoma, leukaemia, allergy
CC   (e.g. asthma), autoimmune disease (e.g. rheumatoid arthritis, lupus,
CC   multiple sclerosis or myasthenia gravis), graft vs. host disease and
CC   transplant or graft rejection.  They inhibit function of interleukin
CC   (IL)-2, -4, -7, -9, -13 and -15 (the receptors for which include a common
CC   gamma -chain), so function as immunosuppressants by reducing
CC   proliferation of T and B cells.
XX
SQ   Sequence   8 AA;


  Query Match             43.4%;  Score 23;  DB 19;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 HDPEF 7
           |:| |
Db       2 heprf 6


RESULT   8
```

```
Y47777
ID   Y47777 standard; Peptide; 9 AA.
XX
AC   Y47777;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2388.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 123; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;


  Query Match             43.4%;  Score 23;  DB 20;  Length 9;
  Best Local Similarity   60.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLHDP 5
            |:| |
Db        3 yvhsp 7


RESULT   9
Y77439
ID   Y77439 standard; peptide; 9 AA.
XX
AC   Y77439;
XX
DT   22-MAY-2000  (first entry)
XX
DE   Fibronectin CS1-derived peptide #30.
XX
KW   Fibronectin; FN; CS-1; endothelial cell; VLA-4 integrin; alpha-4-beta-1;
KW   CD49d/CD29; leukocyte; inflammatory cell; inflammation; cell adhesion;
KW   inhibitor; peptidomimetic; autoimmune disease; inflammatory disorder.
XX
OS   Mammalia.
XX
PN   WO200002903-A1.
XX
PD   20-JAN-2000.
XX
PF   15-DEC-1998;   98WO-US26605.
XX
PR   10-JUL-1998;   98US-0113689.
XX
PA   (CYTE-) CYTEL CORP.
XX
PI   Arrhenius TS,  Elices MJ,  Gaeta FCA,  He Y,  Huyghe BG,  Chen PG;
XX
DR   WPI; 2000-182213/16.
XX
PT   New peptidomimetic compounds used as cell surface fibronectin
PT   expressing receptor and VLA-4 inhibitors for treating inflammatory and
PT   cardiovascular disorders  -
XX
PS   Disclosure; Fig 1; 243pp; English.
XX
CC   The invention relates to peptidomimetic compounds (Y77415-Y77438) capable
CC   of inhibiting the binding of the VLA-4 integrin (alpha-4-beta-1,
CC   CD49d/CD29) to the CS-1 portion (25 amino acids) of a splice variant of
CC   the extracellular matrix protein fibronectin (FN). VLA-4 is expressed on
CC   the surface of leukocytes; the CS-1 FN/VLA-4 interaction plays an
CC   important role in the maturation and trafficking. VLA-4-mediated
CC   leukocyte adhesion to the CS-1 FN of endothelial cells is also a critical
CC   step in the inflammatory response. The peptidomimetics of the invention
CC   may be used to treat both chronic and acute immunoinflammatory
CC   conditions, such as asthma, rheumatoid arthritis, osteoarthritis and
CC   allograft rejection. They may also be used to treat psoriasis and other
CC   skin inflammations, demyelinating diseases of the central nervous system
CC   (e.g., multiple sclerosis), allergies, atherosclerosis, colitis,
CC   diabetes, inflammatory bowel disease, kidney inflammation and restenosis.
CC   Prior art inhibition of VLA-4/CS-1 interaction either involves the use of
CC   anti-VLA-4 antibodies, which can themselves induce an immune response on
CC   repeated administration, or the 25-mer CS-1 peptide, which is large and
CC   costly to make and is subject to rapid proteolytic degradation. The
CC   peptidomimetics of the invention are smaller in comparison to the CS-1
CC   peptide and therefore less expensive to manufacture, and are resistant to
CC   proteolysis. Sequences Y77411-Y77414 and Y77434-Y77444 represent
CC   fragments of the CS-1 peptide tested for their ability to inhibit VLA-4
CC   Jurkat cells to immobilised CS-1 peptide (Y77410).
XX
SQ   Sequence   9 AA;


  Query Match             43.4%;  Score 23;  DB 21;  Length 9;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        2 LHDPE 6
            || ||
Db        1 lhgpe 5


RESULT  10
Y77440
ID   Y77440 standard; peptide; 9 AA.
```

```
XX
AC   Y77440;
XX
DT   22-MAY-2000  (first entry)
XX
DE   Fibronectin CS1-derived peptide #31.
XX
KW   Fibronectin; FN; CS-1; endothelial cell; VLA-4 integrin; alpha-4-beta-1;
KW   CD49d/CD29; leukocyte; inflammatory cell; inflammation; cell adhesion;
KW   inhibitor; peptidomimetic; autoimmune disease; inflammatory disorder.
XX
OS   Mammalia.
XX
PN   WO200002903-A1.
XX
PD   20-JAN-2000.
XX
PF   15-DEC-1998;   98WO-US26605.
XX
PR   10-JUL-1998;   98US-0113689.
XX
PA   (CYTE-) CYTEL CORP.
XX
PI   Arrhenius TS,  Elices MJ,  Gaeta FCA,  He Y,  Huyghe BG,  Chen PG;
XX
DR   WPI; 2000-182213/16.
XX
PT   New peptidomimetic compounds used as cell surface fibronectin
PT   expressing receptor and VLA-4 inhibitors for treating inflammatory and
PT   cardiovascular disorders  -
XX
PS   Disclosure; Fig 1; 243pp; English.
XX
CC   The invention relates to peptidomimetic compounds (Y77415-Y77438) capable
CC   of inhibiting the binding of the VLA-4 integrin (alpha-4-beta-1,
CC   CD49d/CD29) to the CS-1 portion (25 amino acids) of a splice variant of
CC   the extracellular matrix protein fibronectin (FN). VLA-4 is expressed on
CC   the surface of leukocytes; the CS-1 FN/VLA-4 interaction plays an
CC   important role in the maturation and trafficking. VLA-4-mediated
CC   leukocyte adhesion to the CS-1 FN of endothelial cells is also a critical
CC   step in the inflammatory response. The peptidomimetics of the invention
CC   may be used to treat both chronic and acute immunoinflammatory
CC   conditions, such as asthma, rheumatoid arthritis, osteoarthritis and
CC   allograft rejection. They may also be used to treat psoriasis and other
CC   skin inflammations, demyelinating diseases of the central nervous system
CC   (e.g., multiple sclerosis), allergies, atherosclerosis, colitis,
CC   diabetes, inflammatory bowel disease, kidney inflammation and restenosis.
CC   Prior art inhibition of VLA-4/CS-1 interaction either involves the use of
CC   anti-VLA-4 antibodies, which can themselves induce an immune response on
CC   repeated administration, or the 25-mer CS-1 peptide, which is large and
CC   costly to make and is subject to rapid proteolytic degradation. The
CC   peptidomimetics of the invention are smaller in comparison to the CS-1
CC   peptide and therefore less expensive to manufacture, and are resistant to
CC   proteolysis. Sequences Y77411-Y77414 and Y77434-Y77444 represent
CC   fragments of the CS-1 peptide tested for their ability to inhibit VLA-4
CC   Jurkat cells to immobilised CS-1 peptide (Y77410).
XX
SQ   Sequence   9 AA;

  Query Match             43.4%;  Score 23;  DB 21;  Length 9;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LHDPE 6
           || ||
Db       2 lhgpe 6

RESULT  11
W52752
ID   W52752 standard; peptide; 4 AA.
XX
```

```
AC   W52752;
XX
DT   02-MAR-1999  (first entry)
XX
DE   Aminomethylcoumarin-substituted tetrapeptide.
XX
KW   Aminomethylcoumarin; fluorescent; interleukin; ICE; caspase;
KW   positional scanning synthetic combinatorial library.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "the N-terminal is acetylated"
FT   Modified-site   4
FT                   /note= "the C-terminal is condensed onto 7-amino-
FT                          4-methyl-coumarin via an amide linkage"
XX
PN   GB2324529-A.
XX
PD   28-OCT-1998.
XX
PF   20-FEB-1998;   98GB-0003559.
XX
PR   21-FEB-1997;   97US-0038656.
XX
PA   (MERI ) MERCK & CO INC.
XX
PI   Chapman KT,  Nicholson D,  Rano T,  Thornberry N;
XX
DR   WPI; 1998-523841/45.
XX
PT   New tetrapeptide coumarin derivatives - useful in combinatorial
PT   libraries to identify substrate specificity of cysteine or serine
PT   proteases e.g. interleukin-1 converting enzyme
XX
PS   Claim 3; Page 41; 49pp; English.
XX
CC   The peptide is a specifically claimed example of new aminomethyl-
CC   coumarin-labelled tetrapeptides of formula Ac-Xaa-Xaa-Xaa-Asp-AMC.
CC   The tetrapeptides are used to generate a fluorescent positional
CC   scanning synthetic combinatorial library for the investigation of
CC   the substrate specificity of cysteine and serine proteases, e.g. in
CC   the analysis of interleukin-1 beta converting enzyme substrate and
CC   other caspase substrates, and to identify inhibitors which may mediate
CC   inflammations. The library preferably comprises a mixture of at least
CC   200 of the tetrapeptides.
XX
SQ   Sequence   4 AA;

  Query Match             41.5%;  Score 22;  DB 19;  Length 4;
  Best Local Similarity   75.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLHD 4
           |:||
Db       1 yvhd 4

RESULT  12
R14525
ID   R14525 standard; Protein; 5 AA.
XX
AC   R14525;
XX
DT   04-FEB-1992  (first entry)
XX
DE   Interleukin-1 beta protease inhibitor.
XX
KW   human precursor IL-1beta; Graves' disease; Hashimoto's disease.
XX
OS   Homo sapiens.
```

```
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /label= OTHER
FT                   /note= "N-terminally protected"
FT   Modified-site   5
FT                   /label= OTHER
FT                   /note= "carries electronegative leaving group"
XX
PN   WO9115577-A.
XX
PD   17-OCT-1991.
XX
PF   04-APR-1991;   91WO-US02339.
XX
PR   04-APR-1990;   90US-0505298.
PR   13-FEB-1991;   91US-0656759.
XX
PA   (IMMU-) IMMUNEX CORP.
XX
PI   Black RA,  Sleath PR,  Kronheim SR;
XX
DR   WPI; 1991-325220/44.
XX
PT   Isolated Interleukin-1-beta protease - used to improve wound healing
PT   and to treat arthritis and auto-immune diseases e.g. insulin
PT   dependent diabetes
XX
PS   Claim 14; Page 38; 46pp; English.
XX
CC   The 5 amino acids correspond to the sequence of residues 112 to 116
CC   of precursor IL-1 beta (see March et al., Nature, 315: 641-647
CC   (1985)). Inhibitors of the invention comprise from 1 to 5 amino
CC   acid residues corresponding to at least a portion of AYVHD. Specific
CC   examples of suitable inhibitors are given including Boc-Asp-CH2F,
CC   Boc-His-Asp-CH2F, Boc-Phe-Asp-CH2F and Ac-Pro-Asp-CH2F.
XX
SQ   Sequence   5 AA;


  Query Match             41.5%;  Score 22;  DB 12;  Length 5;
  Best Local Similarity   75.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YLHD 4
          |:||
Db      2 yvhd 5


RESULT  13
R34850
ID   R34850 standard; Protein; 5 AA.
XX
AC   R34850;
XX
DT   14-JUL-1993  (first entry)
XX
DE   Interleukin 1-beta protease activity inhibiting peptide.
XX
KW   IL-1beta; therapeutic; wound healing; treatment; arthritis;
KW   autoimmune disease; radiation; side effects.
XX
OS   Synthetic.
XX
PN   WO9305071-A.
XX
PD   18-MAR-1993.
XX
PF   12-SEP-1991;   91WO-US06595.
XX
PR   30-AUG-1991;   91US-0750644.
XX
PA   (IMMV ) IMMUNEX CORP.
```

```
XX
PI   Black RA,  Kronheim SR,  Sleath PR;
XX
DR   WPI; 1993-100924/12.
XX
PT   New polypeptide having protease biological activity for
PT   interleukin 1-beta - can form inhibitor compsn. and cpds. for
PT   treatment of arthritis, auto:immune diseases, inflammation and
PT   radiation damage; also for wound healing
XX
PS   Claim 10; Page 60; 69pp; English.
XX
CC   The sequence is that of pre-interleukin (IL)-1beta amino acids
CC   112-116 which can be used as part of a method of inhibiting
CC   IL-1beta protease activity in a mammal. It can be used in cpds.
CC   for the treatment of inflammation or to treat or prevent
CC   autoimmune disease such as inflammatory disorders.
XX
SQ   Sequence   5 AA;

  Query Match             41.5%;  Score 22;  DB 14;  Length 5;
  Best Local Similarity   75.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YLHD 4
          |:||
Db      2 yvhd 5


RESULT  14
Y21732
ID   Y21732 standard; peptide; 5 AA.
XX
AC   Y21732;
XX
DT   10-SEP-1999  (first entry)
XX
DE   Caspase sensitive site.
XX
KW   Rev-caspase; cysteine protease; zymogen; caspase; cancer; gene therapy;
KW   autoimmune disease; caspase-mediated apoptosis; neurodegenerative;
KW   tumour cell; myocardial infarction; human.
XX
OS   Synthetic.
XX
PN   WO9935277-A2.
XX
PD   15-JUL-1999.
XX
PF   11-JAN-1999;   99WO-US00632.
XX
PR   09-JAN-1998;   98US-0070987.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Alnemri ES;
XX
DR   WPI; 1999-419353/35.
XX
PT   New isolated nucleic acid molecule encoding a rev-caspase  - used
PT   for screening and identifying inhibitors or enhancers for treating
PT   cancer or autoimmune disease
XX
PS   Disclosure; Page 16; 74pp; English.
XX
CC   The invention relates to nucleic acid molecules encoding rev-caspases.
CC   Rev-caspases are cysteine proteases that specifically cleave proteins
CC   after Asp residues and is expressed as a zymogen, in which the small
CC   subunit is N-terminal to a large subunit. A gene delivery vehicle
CC   comprising a rev-caspase coding sequence is useful for the treatment of
CC   cancer, where the gene delivery vehicle is internalised by tumour cells.
CC   The gene delivery vehicle can also be used to treat autoimmune diseases.
```

```
CC   Cells transfected with a rev-caspase expressing vector can be used in
CC   identification of inhibitors or enhancers of caspase-mediated apoptosis.
CC   In vitro translated rev-caspase can be used to identify an inhibitor or
CC   enhancer of caspase processing activity. Caspase inhibitors are useful
CC   for treating neurodegenerative diseases as well as for inhibiting
CC   apoptosis in the heart following myocardial infarction. Sequences Y21728
CC   -Y21733 represent caspase sensitive sites.
XX
SQ   Sequence   5 AA;

  Query Match             41.5%;  Score 22;  DB 20;  Length 5;
  Best Local Similarity   75.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLHD 4
            |:||
Db        1 yvhd 4


RESULT  15
P93063
ID   P93063 standard; peptide; 6 AA.
XX
AC   P93063;
XX
DT   12-MAR-1990  (first entry)
XX
DE   Peptide encoded by degenerate probe for soybean kunitz trypsin inhibitor.
XX
KW   Probe; soybean; kunitz-type trypsin inhibitor.
XX
OS   Glycine max.
XX
PN   EP339009-A.
XX
PD   25-OCT-1989.
XX
PF   10-APR-1989;   89EP-0870047.
XX
PR   11-APR-1988;   88US-0179709.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Fuchs RL,  Kishore GM,  MacIntosh SC;
XX
DR   WPI; 1989-311431/43.
DR   N-PSDB; N91714.
XX
PT   Toxin protein of Bacillus thuringiensis bacteria - improved in efficacy
PT   using a potentiating amt. of a trypsin inhibitor.  The sequence contains
PT   IUPAC ambiguity codes.
XX
PS   Disclosure; page 8; 56pp; English.
XX
CC   The peptide is encoded by a degenerate probe and corresponds to amino
CC   acids 125-130 of soybean kunitz trypsin inhibitor.
XX
SQ   Sequence   6 AA;

  Query Match             41.5%;  Score 22;  DB 10;  Length 6;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 DPEFN 8
            | |||
Db        1 ddefn 5


RESULT  16
R34855
ID   R34855 standard; Protein; 8 AA.
```

```
XX
AC   R34855;
XX
DT   14-JUL-1993  (first entry)
XX
DE   IL-1beta protease cleavage site peptide substrate.
XX
KW   Interleukin; therapeutic; wound healing; treatment; arthritis;
KW   autoimmune disease; radiation; side effects.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Cleavage-site   3..6
FT                   /note= "IL-1beta protease cleavage site"
XX
PN   WO9305071-A.
XX
PD   18-MAR-1993.
XX
PF   12-SEP-1991;   91WO-US06595.
XX
PR   30-AUG-1991;   91US-0750644.
XX
PA   (IMMV ) IMMUNEX CORP.
XX
PI   Black RA,  Kronheim SR,  Sleath PR;
XX
DR   WPI; 1993-100924/12.
XX
PT   New polypeptide having protease biological activity for
PT   interleukin 1-beta - can form inhibitor compsn. and cpds. for
PT   treatment of arthritis, auto:immune diseases, inflammation and
PT   radiation damage; also for wound healing
XX
PS   Example; Page 39; 69pp; English.
XX
CC   The sequence is that of a peptide corresponding to Tyr113 to Arg120
CC   of human precursor interleukin (IL)-1beta sequence which contains
CC   the cleavage site (His115-Pro118) for IL-1beta protease. It shows
CC   a reactivity of 2.40 relative to the peptide comprising amino acids
CC   Ala112-Ser121 (R34851).
XX
SQ   Sequence   8 AA;

  Query Match             41.5%;  Score 22;  DB 14;  Length 8;
  Best Local Similarity   75.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLHD 4
            |:||
Db        1 yvhd 4


RESULT  17
R12720
ID   R12720 standard; Protein; 5 AA.
XX
AC   R12720;
XX
DT   31-JUL-1991  (first entry)
XX
DE   Pentapeptide paralogue for use as chromatographic affinity ligand.
XX
KW   HPLC.
XX
OS   Synthetic.
XX
PN   WO9106356-A.
XX
PD   16-MAY-1991.
XX
```

```
PF   31-OCT-1990;   90WO-EP06333.
XX
PR   31-OCT-1989;   89US-0429721.
XX
PA   (TERR-) TERRAPIN TECHN INC.
XX
PI   Kauvar LM;
XX
DR   WPI; 1991-163985/22.
XX
PT   Identifying paralog with specific affinity for analyte - using
PT   candidate paralogs with systematically varied values of at least
PT   2 parameters.
XX
PS   Claim 41; Fig 8; 95pp; English.
XX
CC   Peptide is one of a panel, each of which have systemicly varied
CC   values of at least two parameters. A test protein is matched against
CC   the panel, and the paralogues with highest binding affinity are
CC   selected. Selected paralogues are bound to a matrix which is
CC   incorporated into the HPLC gel and used as affinity ligands.
CC   Technique is useful in chromatographic separation, purification
CC   and binding assay.
XX
SQ   Sequence   5 AA;


  Query Match             39.6%;  Score 21;  DB 12;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 HDP 5
            |||
Db        2 hdp 4


RESULT  18
R51584
ID   R51584 standard; peptide; 5 AA.
XX
AC   R51584;
XX
DT   19-OCT-1994  (first entry)
XX
DE   Mimotope peptide #78 from panel of maximally diverse mimotopes.
XX
KW   mimotope panel; rational drug design; candidate drug;
KW   screening assay; hydrophobicity; antibody repertoire.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "acetylated"
XX
PN   US5300425-A.
XX
PD   05-APR-1994.
XX
PF   13-OCT-1987;   87US-0108130.
XX
PR   13-OCT-1987;   87US-0108130.
PR   11-OCT-1988;   88US-0255906.
PR   06-DEC-1989;   89US-0447009.
XX
PA   (TERR-) TERRAPIN TECHNOLOGIES INC.
XX
PI   Kauvar LM;
XX
DR   WPI; 1994-109390/13.
XX
PT   Screening of candidate drugs for binding to receptor - by
PT   comparing inverse image antibody profile of drug with
PT   mimotype-binding profile of receptor
XX
PS   Example 3; Fig 3; 29pp; English.
XX
CC   A panel of 88 pentapeptides was designed on the basis of decreasing
CC   hydrophobicity and periodic variation of hydrophobic moment. The
CC   peptides were labelled with iodine-125 and tested with individual
CC   members of a basal antibody repertoire. (Spleen cells were
CC   harvested from mice and used to provide a panel of
CC   antibody-secreting hybridoma cells as a subset of the complete
CC   B-cell repertoire). Nearly uniform binding to all antibody members
CC   of the repertoire was observed. The test was than repeated with the
CC   addition of a defined amount of analyte to the mixture. A small
CC   number of wells showed greatly decreased labelling and these
CC   antibodies represented the successful result of an initial screen
CC   for those which preferentially bind analyte.
XX
SQ   Sequence   5 AA;


  Query Match             39.6%;  Score 21;  DB 15;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 HDP 5
            |||
Db        2 hdp 4


RESULT  19
R69952
ID   R69952 standard; peptide; 5 AA.
XX
AC   R69952;
XX
DT   19-OCT-1995  (first entry)
XX
DE   Pentameric mimotope 78 used to obtain highly specific antibodies.
XX
KW   mimotope; antibody; production; high specificity; detection;
KW   immunoassay; high performance liquid chromatography.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "Acetyl-Ser"
XX
PN   US5384263-A.
XX
PD   24-JAN-1995.
XX
PF   13-OCT-1987;   87US-0108130.
XX
PR   11-OCT-1988;   88US-0255906.
PR   13-OCT-1987;   87US-0108130.
PR   04-JUN-1993;   93US-0072190.
XX
PA   (TERR-) TERRAPIN TECHNOLOGIES INC.
XX
PI   Kauvar LM;
XX
DR   WPI; 1995-105497/14.
XX
PT   Producing antibodies with high specificity and affinity for an
PT   analyte  - by immunisation with selected mimotope, also analyte
PT   detection kits, useful for immunoassay of materials usually
PT   analysed by HPLC
XX
PS   Example 3; Fig 3; 25pp; English.
XX
CC   R69875-969 are pentameric mimotopes designed on the basis of
CC   decreasing hydrophobicity and periodic variation of hydrophobic
```

```
CC   moment. All the pentapeptides are acetylated at the N-terminus except
CC   E12 (for FITC) (sic). E12 is not identified in the specification.
CC   The mimotopes are used in the method of the invention to obtain
CC   antibodies specifically and strongly reactive  with a desired
CC   analyte. The mimotope is obtd. by reacting a panel of starting
CC   antibodies (Abs) representative of the resting B cell repertoire of a
CC   mammal with an analyte (so as to identify analyte-reacting Abs) and
CC   then reacting each of a panel of candidate mimotopes representative
CC   of a random set of 3D contours with the analyte-reacting Abs. A
CC   subject is immunised with one or more mimotopes identified and the
CC   product Abs are recovered from the serum of the subject.
XX
SQ   Sequence   5 AA;

  Query Match             39.6%;  Score 21;  DB 16;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HDP 5
              |||
Db          2 hdp 4


RESULT  20
R98698
ID   R98698 standard; peptide; 5 AA.
XX
AC   R98698;
XX
DT   11-MAR-1997  (first entry)
XX
DE   Peptide 78 from 88 member diverse mimotope panel.
XX
KW   Panel; mimotope; decreasing hydrophobicity; periodic variation;
KW   hydrophobic moment; antibody; repertoire; identification; drug;
KW   candidate; receptor; binding; ligand; rational; design; selection;
KW   treatment; tumour; production; immunological reagent; analyte;
KW   detection; trace contaminant; mimetope.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "acylated"
XX
PN   US5541070-A.
XX
PD   30-JUL-1996.
XX
PF   13-OCT-1987;   87US-0108130.
XX
PR   06-DEC-1989;   89US-0447009.
PR   13-OCT-1987;   87US-0108130.
PR   11-OCT-1988;   88US-0255906.
PR   08-SEP-1993;   93US-0118133.
XX
PA   (KAUV/) KAUVAR L M.
XX
PI   Kauvar LM;
XX
DR   WPI; 1996-361955/36.
XX
PT   Identifying candidate drugs that bind a specific receptor by
PT   competitive reaction with panel of mimetope(s) - useful in rational
PT   drug design
XX
PS   Example 3; Fig 3; 27pp; English.
XX
CC   The present peptide is a member of a panel of 88 pentapeptide
CC   mimotopes designed on the basis of decreasing hydrophobicity, and
CC   periodic variation of hydrophobic moment. The panel was synthesised
CC   using the method of Geysen, H. M., et al, Proc. Natl. Acad. Sci.
```

```
CC   USA (1984), which uses lots of 96 pins; the remaining 8
CC   polyethylene pins being controls. The mimotopes were then mixed,
CC   125-I labelled and tested with individual members of a basal
CC   antibody (Ab) repertoire. Nearly uniform binding to all members was
CC   found. The test was then repeated with the addition of a defined
CC   amt. of analyte. A small number showed greatly increased labelling,
CC   these Ab representing the successful result of an initial screen
CC   for those that pref. bind analyte.
CC   The above is an example of a claimed method for identifying members
CC   of a panel of candidate drugs, that bind to a receptor having a
CC   known ligand. It is useful in rational drug design, e.g. selection
CC   of monoclonal Ab for treating individual tumours, and for the prodn.
CC   of immunological reagents for any analyte, including those not
CC   normally detectable by immunoassay, e.g trace contaminants in soil,
CC   air or water.
XX
SQ   Sequence   5 AA;

  Query Match             39.6%;  Score 21;  DB 17;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HDP 5
              |||
Db          2 hdp 4


RESULT  21
Y23613
ID   Y23613 standard; Peptide; 5 AA.
XX
AC   Y23613;
XX
DT   03-SEP-1999  (first entry)
XX
DE   SEQ ID 104 of US5919900.
XX
KW   Neutrophil inhibitory factor; NIF; mutant; shock; stroke;
KW   allograft rejection; vasculitis; autoimmune diabetes; ARDS;
KW   rheumatoid arthritis; inflammatory skin disease; myocardial infarction;
KW   inflammatory bowel disease; adult respiratory distress syndrome;
KW   ischemia-reperfusion injury; acute inflammation; bacterial infection;
KW   vaccine; parasitic worm infection; antihelminic.
XX
OS   Undentified.
XX
PN   US5919900-A.
XX
PD   06-JUL-1999.
XX
PF   26-MAY-1995;   95US-0450497.
XX
PR   23-DEC-1993;   93US-0173510.
PR   11-MAY-1992;   92US-0881721.
PR   24-DEC-1992;   92US-0996972.
PR   11-MAY-1993;   93US-0060433.
PR   10-NOV-1993;   93US-0151064.
PR   26-MAY-1995;   95US-0450497.
XX
PA   (CORV-) CORVAS INT INC.
XX
PI   Foster DL,  Moyle M;
XX
DR   WPI; 1999-403975/34.
XX
PT   Mutant Neutrophil Inhibitory Factors useful for treating
PT   inflammatory conditions and especially to prevent or decrease
PT   inflammatory responses
XX
PS   Disclosure; Column 151; 131pp; English.
XX
CC   The specification describes mutant Neutrophil Inhibitory Factors (NIFs),
```

```
CC   where at least 1 of the Asp residues at positions 18, 87, 110, 130, 197,
CC   or 223 in the wild type sequence (see Y23591) is replaced by a Gln
CC   residue. NIFs may be useful for treating shock, stroke, acute and
CC   chronic allograft rejection, vasculitis, autoimmune diabetes, rheumatoid
CC   arthritis, inflammatory skin diseases, inflammatory bowel disease, adult
CC   respiratory distress syndrome (ARDS), ischemia-reperfusion injury
CC   following myocardial infarction, and acute inflammation caused by
CC   bacterial infection such as sepsis or bacterial meningitis. NIFs or NIF
CC   fragments may be used as vaccines against parasitic worm infection.
CC   Anti-NIF antibodies may be useful for detecting infection of a mammalian
CC   host by a parasitic worm, as antihelminic agents, and in the detection
CC   and isolation of NIF from tissue homogenates, cloned cells etc. NIFs may
CC   be useful for the detection of NIF mimics or antagonists in other
CC   compounds. Other NIF agonists and inhibitors may also be used as
CC   antihelminic agents. The present sequence is used in the course of
CC   the invention.
XX
SQ   Sequence   5 AA;


  Query Match             39.6%;  Score 21;  DB 20;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         3 HDP 5
             |||
Db         3 hdp 5


RESULT  22
R20337
ID   R20337 standard; Protein; 7 AA.
XX
AC   R20337;
XX
DT   19-APR-1992  (first entry)
XX
DE   Sequence of synthetic peptide based on clostripain digest fragment
DE   of Cathepsin G (CAT. G).
XX
KW   Antimicrobial; antibacterial; gingivitis; periodontosis; therapy;
KW   tumouricide; virucide; anti-viral agent; fungicide;
KW   enzyme; protease.
XX
OS   Homo sapiens.
XX
PN   WO9119512-A.
XX
PD   26-DEC-1991.
XX
PF   21-JUN-1991;   91WO-US04414.
XX
PR   21-JUN-1990;   90US-0541635.
XX
PA   (UYGE-) UNIV GEORGIA RES.
PA   (UYEM-) EMORY UNIV.
XX
PI   Travis J,  Shafer WM,  Bangalore N,  Pohl J;
XX
DR   WPI; 1992-024198/03.
XX
PT   New oligopeptide(s) with antimicrobial and antibacterial activity
PT   - useful in treatment of bacterial infection, e.g. gingivitis or
PT   periodontosis and to kill or control growth of tumour or
PT   virus-infected cells
XX
PS   Claim 10; Page 41; 54pp; English.
XX
CC   Two of the peptides of the invention (R20331 and R20333) are the
CC   products of the digestion of human Cathepsin G (Cat. G) with
CC   clostripain. Similar peptides were then synthesised and tested for
CC   antimicrobial activity. 20 peptides, including R20331 and
CC   R20333, are claimed. The peptides are active against Gram-
```

```
CC   positive and negative bacteria, fungi and protozoa. They can be used
CC   for treatment or prevention of infections and for the treatment and/
CC   or prevention of gingivitis. The IIGGR-related (R20331, R20332) can
CC   also be used to kill or control the growth of tumour cells or virus-
CC   infected cells.
XX
SQ   Sequence   7 AA;

  Query Match             39.6%;  Score 21;  DB 13;  Length 7;
  Best Local Similarity   75.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         5 PEFN 8
             |:||
Db         2 pqfn 5


RESULT  23
R20345
ID   R20345 standard; Protein; 7 AA.
XX
AC   R20345;
XX
DT   19-APR-1992  (first entry)
XX
DE   Sequence of synthetic peptide based on clostripain digest fragment
DE   of Cathepsin G (CAT. G).
XX
KW   Antimicrobial; antibacterial; gingivitis; periodontosis; therapy;
KW   tumouricide; virucide; anti-viral agent; fungicide;
KW   enzyme; protease.
XX
OS   Homo sapiens.
XX
PN   WO9119512-A.
XX
PD   26-DEC-1991.
XX
PF   21-JUN-1991;   91WO-US04414.
XX
PR   21-JUN-1990;   90US-0541635.
XX
PA   (UYGE-) UNIV GEORGIA RES.
PA   (UYEM-) EMORY UNIV.
XX
PI   Travis J,  Shafer WM,  Bangalore N,  Pohl J;
XX
DR   WPI; 1992-024198/03.
XX
PT   New oligopeptide(s) with antimicrobial and antibacterial activity
PT   - useful in treatment of bacterial infection, e.g. gingivitis or
PT   periodontosis and to kill or control growth of tumour or
PT   virus-infected cells
XX
PS   Claim 10; Page 41; 54pp; English.
XX
CC   Two of the peptides of the invention (R20331 and R20333) are the
CC   products of the digestion of human Cathepsin G (Cat. G) with
CC   clostripain. Similar peptides were then synthesised and tested for
CC   antimicrobial activity. 20 peptides, including R20331 and
CC   R20333, are claimed. The peptides are active against Gram-
CC   positive and negative bacteria, fungi and protozoa. They can be used
CC   for treatment or prevention of infections and for the treatment and/
CC   or prevention of gingivitis. The IIGGR-related (R20331, R20332) can
CC   also be used to kill or control the growth of tumour cells or virus-
CC   infected cells.
XX
SQ   Sequence   7 AA;

  Query Match             39.6%;  Score 21;  DB 13;  Length 7;
  Best Local Similarity   75.0%;  Pred. No. 2.1e+05;
```

```
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 PEFN 8
           ||:|
Db       2 peyn 5


RESULT  24
R55584
ID   R55584 standard; Peptide; 7 AA.
XX
AC   R55584;
XX
DT   18-JAN-1995  (first entry)
XX
DE   AmEPV spheroidin probe RM58 encoded peptide.
XX
KW   Spheroidin; gene expression; vector; insect cell culture;
KW   mammal cell culture; AmEPV; Amsacta moorei; entomopoxvirus;
KW   DNA probe; hybridization.
XX
OS   Synthetic.
XX
PN   WO9413812-A.
XX
PD   23-JUN-1994.
XX
PF   07-DEC-1993;   93WO-US11907.
XX
PR   07-DEC-1992;   92US-0991867.
XX
PA   (UYFL ) UNIV FLORIDA.
XX
PI   Gruidl ME,  Hall RL,  Moyer RW;
XX
DR   WPI; 1994-217887/26.
DR   N-PSDB; Q66799.
XX
PT   New entomopoxvirus polynucleotide sequences, proteins and vectors
PT   - are used for expression of heterologous proteins in both insect
PT   and mammalian host cells
XX
PS   Disclosure; Page 84; 118pp; English.
XX
CC   Degenerate oligonucleotide RM58 was used as a hybridization
CC   probe to locate the spheroidin gene in an AmEPV DNA library.
CC   The peptide encoded by the probe is given in R55584.
XX
SQ   Sequence   7 AA;

  Query Match             39.6%;  Score 21;  DB 15;  Length 7;
  Best Local Similarity   75.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 DPEF 7
           |||:
Db       3 dpey 6


RESULT  25
Y17008
ID   Y17008 standard; peptide; 7 AA.
XX
AC   Y17008;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Heat shock protein (hsp) binding peptide.
XX
KW   Conjugate peptide; heat shock protein; hsp; phage display library; virus;
KW   surface protein; tethering peptide; chaperone process; cytokine; cancer;
KW   neoplastic disease; infectious disease; bacterium; immune system; fungus;
```

```
KW   acquired immune deficiency; autoimmune disease.
XX
OS   Synthetic.
XX
PN   WO9922761-A1.
XX
PD   14-MAY-1999.
XX
PF   22-OCT-1998;   98WO-US22335.
XX
PR   31-OCT-1997;   97US-0961707.
XX
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI   Hartl U,  Hoe MH,  Houghton A,  Mayhew M,  Moroi Y;
PI   Ouerfelli O,  Rothman JE;
XX
DR   WPI; 1999-313177/26.
XX
PT   Identifying peptides which bind heat shock proteins
XX
PS   Examples; Page 24; 155pp; English.
XX
CC   The invention relates to conjugate peptides engineered to noncovalently
CC   bind to heat shock proteins (hsp). A method of identifying a hsp binding
CC   peptide comprises (a) contacting a phage display library having
CC   bacteriophage expressing, in a surface protein, inserted peptides with a
CC   hsp target, and bound to a benzoquinone ansamycin antibiotic (BAA), in a
CC   physiologic binding buffer; (b) isolating a phage binding to the hsp
CC   target; and (c) identifying the inserted peptide expressed. The peptides
CC   which bind to a hsp can be used as tethering peptides for a hsp which may
CC   serve as an accessory in a chaperone process and/or may comprise a
CC   cytokine. They can also be coupled to antigens to induce an immune
CC   response. Such compositions can be used for treating neoplastic disease,
CC   e.g. cancers, infectious diseases, e.g. diseases caused by a bacterium,
CC   virus, protozoan, mycoplasma, fungus, yeast, parasite or prion, or a
CC   disease of the immune system, e.g. acquired immune deficiencies or
CC   autoimmune diseases.
XX
SQ   Sequence   7 AA;

  Query Match             39.6%;  Score 21;  DB 20;  Length 7;
  Best Local Similarity   60.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLHDP 5
            :|| |
Db       1 hlhsp 5


RESULT  26
Y16837
ID   Y16837 standard; peptide; 8 AA.
XX
AC   Y16837;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Heat shock protein (hsp) binding peptide.
XX
KW   Conjugate peptide; heat shock protein; hsp; phage display library; virus;
KW   surface protein; tethering peptide; chaperone process; cytokine; cancer;
KW   neoplastic disease; infectious disease; bacterium; immune system; fungus;
KW   acquired immune deficiency; autoimmune disease.
XX
OS   Synthetic.
XX
PN   WO9922761-A1.
XX
PD   14-MAY-1999.
XX
PF   22-OCT-1998;   98WO-US22335.
```

```
XX
PR   31-OCT-1997;    97US-0961707.
XX
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI   Hartl U,  Hoe MH,  Houghton A,  Mayhew M,  Moroi Y;
PI   Ouerfelli O,  Rothman JE;
XX
DR   WPI; 1999-313177/26.
DR   N-PSDB; X60737.
XX
PT   Identifying peptides which bind heat shock proteins
XX
PS   Disclosure; Fig 1H; 155pp; English.
XX
CC   The invention relates to conjugate peptides engineered to noncovalently
CC   bind to heat shock proteins (hsp). A method of identifying a hsp binding
CC   peptide comprises (a) contacting a phage display library having
CC   bacteriophage expressing, in a surface protein, inserted peptides with a
CC   hsp target, and bound to a benzoquinone ansamycin antibiotic (BAA), in a
CC   physiologic binding buffer; (b) isolating a phage binding to the hsp
CC   target; and (c) identifying the inserted peptide expressed. The peptides
CC   which bind to a hsp can be used as tethering peptides for a hsp which may
CC   serve as an accessory in a chaperone process and/or may comprise a
CC   cytokine. They can also be coupled to antigens to induce an immune
CC   response. Such compositions can be used for treating neoplastic disease,
CC   e.g. cancers, infectious diseases, e.g. diseases caused by a bacterium,
CC   virus, protozoan, mycoplasma, fungus, yeast, parasite or prion, or a
CC   disease of the immune system, e.g. acquired immune deficiencies or
CC   autoimmune diseases.
XX
SQ   Sequence   8 AA;


  Query Match             39.6%;  Score 21;  DB 20;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLHDP 5
            :|| |
Db        1 hlhsp 5


RESULT  27
R67767
ID   R67767 standard; Protein; 9 AA.
XX
AC   R67767;
XX
DT   23-FEB-1995  (first entry)
XX
DE   N-terminal sequence of neutrophil inhibitory factor (AcaNIF9).
XX
KW   Primer; neutrophil inhibitory factor; NIF; inflammation; adhesion;
KW   endothelial cells; inflammatory response.
XX
OS   Ancylostoma caninum.
XX
PN   WO9414973-A.
XX
PD   07-JUL-1994.
XX
PF   23-DEC-1993;    93WO-US12626.
XX
PR   24-DEC-1992;    92US-0996972.
PR   11-MAY-1993;    93US-0060433.
PR   10-NOV-1993;    93US-0151064.
XX
PA   (CORV-) CORVAS INT INC.
XX
PI   Foster DL,  Moyle M,  Vlasuk GP;
XX
DR   WPI; 1994-234706/28.
```

```
XX
PT   Neutrophil inhibitory factor peptide(s) - derived from nematodes,
PT   useful for therapy of inflammatory responses
XX
PS   Example 22(C); Page 125; 239pp; English.
XX
CC   Neutrophil inhibitory factors can be used in compositions to inhibit
CC   neutrophil activity e.g. adhesion to vascular endothelial cells, and
CC   which are useful in the therapy of conditions which involve abnormal
CC   or undesired inflammatory responses.  The N-terminal end of the
CC   proteolytically processed NIF clone designated AceNIF3 is ambiguous
CC   and so for cloning purposes an N-terminal sequence was generated
CC   based on sequence homology between this NIF clone and AceNIF3.
CC   (R67768).
XX
SQ   Sequence   9 AA;


  Query Match             39.6%;  Score 21;  DB 15;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      . 3 HDP 5
            |||
Db        3 hdp 5


RESULT  28
Y48130
ID   Y48130 standard; Peptide; 9 AA.
XX
AC   Y48130;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2741.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;    98WO-US05039.
XX
PR   13-MAR-1998;    98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 138; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
```

```
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;


  Query Match             39.6%;  Score 21;  DB 20;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 HDP 5
            |||
Db        4 hdp 6


RESULT  29
Y23570
ID   Y23570 standard; Peptide; 9 AA.
XX
AC   Y23570;
XX
DT   03-SEP-1999  (first entry)
XX
DE   N-terminal sequene of a hookworm NIF protein.
XX
KW   Neutrophil inhibitory factor; NIF; mutant; shock; stroke;
KW   allograft rejection; vasculitis; autoimmune diabetes; ARDS;
KW   rheumatoid arthritis; inflammatory skin disease; myocardial infarction;
KW   inflammatory bowel disease; adult respiratory distress syndrome;
KW   ischemia-reperfusion injury; acute inflammation; bacterial infection;
KW   vaccine; parasitic worm infection; antihelminic.
XX
OS   Ancylostoma ceylanicum.
XX
PN   US5919900-A.
XX
PD   06-JUL-1999.
XX
PF   26-MAY-1995;   95US-0450497.
XX
PR   23-DEC-1993;   93US-0173510.
PR   11-MAY-1992;   92US-0881721.
PR   24-DEC-1992;   92US-0996972.
PR   11-MAY-1993;   93US-0060433.
PR   10-NOV-1993;   93US-0151064.
PR   26-MAY-1995;   95US-0450497.
XX
PA   (CORV-) CORVAS INT INC.
XX
PI   Foster DL,  Moyle M;
XX
DR   WPI; 1999-403975/34.
XX
PT   Mutant Neutrophil Inhibitory Factors useful for treating
PT   inflammatory conditions and especially to prevent or decrease
PT   inflammatory responses
XX
PS   Example 22; Column 63; 131pp; English.
XX
CC   The specification describes mutant Neutrophil Inhibitory Factors (NIFs),
CC   where at least 1 of the Asp residues at positions 18, 87, 110, 130, 197,
CC   or 223 in the wild type sequence (see Y23591) is replaced by a Gln
```

```
CC   residue. NIFs may be useful for treating shock, stroke, acute and
CC   chronic allograft rejection, vasculitis, autoimmune diabetes, rheumatoid
CC   arthritis, inflammatory skin diseases, inflammatory bowel disease, adult
CC   respiratory distress syndrome (ARDS), ischemia-reperfusion injury
CC   following myocardial infarction, and acute inflammation caused by
CC   bacterial infection such as sepsis or bacterial meningitis, NIFs or NIF
CC   fragments may be used as vaccines against parasitic worm infection.
CC   Anti-NIF antibodies may be useful for detecting infection of a mammalian
CC   host by a parasitic worm, as antihelminic agents, and in the detection
CC   and isolation of NIF from tissue homogenates, cloned cells etc. NIFs may
CC   be useful for the detection of NIF mimics or antagonists in other
CC   compounds. Other NIF agonists and inhibitors may also be used as
CC   antihelminic agents. The present sequence represents a N-terminal
CC   of hookworm NIF.
XX
SQ   Sequence   9 AA;


  Query Match             39.6%;  Score 21;  DB 20;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 HDP 5
            |||
Db        3 hdp 5


RESULT  30
W89129
ID   W89129 standard; peptide; 9 AA.
XX
AC   W89129;
XX
DT   10-MAR-1999  (first entry)
XX
DE   ScFV (dig) variant clone LC2-7 isolated by FACS.
XX
KW   Enzyme; antibody; catalysing; immune response; digoxin; digitoxin; scFv;
KW   assay; cytokine; transcription factor; clotting factor; chelating agent;
KW   hormone; receptor; single chain Fv; flourescence activated cell sorting;
KW   FACS; variant.
XX
OS   Unidentified.
XX
PN   WO9849286-A2.
XX
PD   05-NOV-1998.
XX
PF   30-APR-1998;   98WO-US08714.
XX
PR   01-MAY-1997;   97US-0847063.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Chen G,  Daugherty PS,  Georgiou G,  Iverson B,  Olsen MJ;
XX
DR   WPI; 1999-024054/02.
XX
PT   Selection of desired polypeptides from candidates - comprises use of
PT   host cells containing a library of vectors that encode distinct
PT   candidate polypeptides which are expressed on the surface of the
PT   cells
XX
PS   Example 7; Page 81; 149pp; English.
XX
CC   The invention relates to methods for selecting a polypeptide from
CC   candidate proteins. The  method comprises obtaining a library of vectors
CC   that encode distinct candidate polypeptides, where the vectors provide
CC   for the cell surface expression of the candidate polypeptides; expressing
CC   each of the candidate polypeptides on the surface of a host cell, and
CC   selecting a host cell that expresses a desired polypeptide. Methods for
CC   catalysing a chemical reaction and methods for stimulating an immune
CC   response are also provided. A host cell  that expresses, on its surface,
```

```
CC   a single-chain antibody that binds immunologically to a digoxin, but does
CC   not bind immunologically to digitoxin is also provided. The methods can
CC   be used for obtaining desired polypeptides such as an antibody or
CC   antibody fragment, an enzyme, a cytokine, a transcription factor, a
CC   clotting factor, a chelating agent, a hormone or a receptor. Assays can
CC   be conducted with remarkable rapidity as the polypeptides are displayed
CC   on the surface of a host cell. The methods allow production of large
CC   quantities of these polypeptides, potentially on a kilogram scale, from
CC   microorganism cultures. Sequences W89128 to W89132 represent anti-digoxin
CC   single chain Fv (scFV) antibody variants isolated from a light chain
CC   library LCDR3 by flourescence activated cell sorting (FACS).
XX
SQ   Sequence   9 AA;


  Query Match             39.6%;  Score 21;  DB 20;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 HDP 5
              |||
Db          5 hdp 7


RESULT  31
W86288
ID   W86288 standard; peptide; 6 AA.
XX
AC   W86288;
XX
DT   19-FEB-1999  (first entry)
XX
DE   Rodent interleukin (IL)-1 delta polypeptide fragment (residues 19-24).
XX
KW   Interleukin; IL-1 delta; polyclonal antibody; IL-1 epsilon; cytokine;
KW   inflammatory response; immune system; diagnosis; agonist; antagonist;
KW   chemokine.
XX
OS   Mus sp.
XX
PN   WO9847921-A1.
XX
PD   29-OCT-1998.
XX
PF   17-APR-1998;   98WO-US06879.
XX
PR   06-AUG-1997;   97US-0055111.
PR   21-APR-1997;   97US-0837627.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Bazan JF,  Hedrick JA,  Kastelein RA,  Sana TR;
XX
DR   WPI; 1998-609976/51.
XX
PT   Mammalian interleukin 1-delta and 1-epsilon - useful for, e.g.
PT   regulating the immune system and inflammatory responses
XX
PS   Claim 1; Page 99; 113pp; English.
XX
CC   The invention relates to a recombinant polypeptide that specifically
CC   binds polyclonal antibodies (Abs) generated against a 12 consecutive
CC   amino acid segment of interleukin (IL)-1 delta or IL-1 epsilon. Agonists
CC   or antagonists of these IL polypeptides are used to regulate a cell
CC   involved in an inflammatory response. The IL-1 delta or IL-1 epsilon
CC   polypeptides and peptides are used to produce Abs and antigen-Ab
CC   complexes. The polypeptides, Abs and the corresponding nucleic acids
CC   regulate development and/or the immune system, and can be used to
CC   diagnose and treat conditions associated with abnormal expression of IL.
CC   Agonists or antagonists of IL-1 delta or IL-1 epsilon polypeptides are
CC   used with agonists or antagonists of IL-1 alpha, IL-1RA, IL-1 beta, IL-1
CC   gamma, IL-2 and/or IL-12. The IL-1 delta or IL-1 epsilon polypeptides may
CC   be used as a soluble polypeptide or as a fusion protein with another
CC   cytokine or chemokine. Sequences W86287 to W86300 represent peptide
CC   fragments of a rodent interleukin (IL)-1 delta polypeptide.
XX
SQ   Sequence   6 AA;


  Query Match             37.7%;  Score 20;  DB 19;  Length 6;
  Best Local Similarity   75.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLHD 4
              |||:
Db          3 ylhn 6


RESULT  32
W87325
ID   W87325 standard; peptide; 6 AA.
XX
AC   W87325;
XX
DT   09-FEB-1999  (first entry)
XX
DE   Peptide determined by the method of the invention.
XX
KW   Amino acid determination; molecular mass; fragmentation spectrum;
KW   DNA cloning; anti-body; recombinant; modification; mass spectrometry.
XX
OS   Synthetic.
XX
PN   GB2325465-A.
XX
PD   25-NOV-1998.
XX
PF   22-MAY-1998;   98GB-0011196.
XX
PR   22-MAY-1997;   97GB-0010582.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Parekh RB,  Prime SB,  Townsend RR,  Wedd NS;
XX
DR   WPI; 1998-571195/49.
XX
PT   Peptide sequence determination used in e.g. DNA cloning - by
PT   comparing mass spectra of the unknown peptide with a library of
PT   linear chain known peptide sequences
XX
PS   Example 1; Page 19; 40pp; English.
XX
CC   The invention relates to a method for determination of the amino acid
CC   sequence of an unknown peptide. The method comprises (a) determining
CC   the molecular mass and an experimental fragmentation spectrum for the
CC   peptide; (b) comparing the experimental fragmentation spectrum of the
CC   unknown peptide with a theoretical fragmentation spectra calculated for
CC   a peptide library composed of all possible linear sequences of amino
CC   acids having a total mass that corresponds to the molecular mass of the
CC   unknown peptide; and (c) identifying a peptide in the library with a
CC   theoretical fragmentation spectrum that most closely matches the
CC   fragmentation spectrum of the unknown peptide. The method is useful in
CC   DNA cloning, anti-body production, identification of recombinant
CC   products, and the study of post-translational modifications. It allows
CC   the sequence of unknown peptides or proteins with no sub-sequence
CC   identity, to be characterised using mass spectrometry. Sequences W87101
CC   to W87364 represent a library of linear peptides constructed to exemplify
CC   the method. The isoleucine residue in these peptides can be replaced by
CC   leucine to construct another 264 linear peptides to be included in the
CC   library.
XX
SQ   Sequence   6 AA;


  Query Match             37.7%;  Score 20;  DB 19;  Length 6;
```

```
  Best Local Similarity   75.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLHD 4
          | ||
Db      1 yahd 4


RESULT  33
W87276
ID   W87276 standard; peptide; 6 AA.
XX
AC   W87276;
XX
DT   09-FEB-1999  (first entry)
XX
DE   Peptide determined by the method of the invention.
XX
KW   Amino acid determination; molecular mass; fragmentation spectrum;
KW   DNA cloning; anti-body; recombinant; modification; mass spectrometry.
XX
OS   Synthetic.
XX
PN   GB2325465-A.
XX
PD   25-NOV-1998.
XX
PF   22-MAY-1998;   98GB-0011196.
XX
PR   22-MAY-1997;   97GB-0010582.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Parekh RB,  Prime SB,  Townsend RR,  Wedd NS;
XX
DR   WPI; 1998-571195/49.
XX
PT   Peptide sequence determination used in e.g. DNA cloning - by
PT   comparing mass spectra of the unknown peptide with a library of
PT   linear chain known peptide sequences
XX
PS   Example 1; Page 20; 40pp; English.
XX
CC   The invention relates to a method for determination of the amino acid
CC   sequence of an unknown peptide. The method comprises (a) determining
CC   the molecular mass and an experimental fragmentation spectrum for the
CC   peptide; (b) comparing the experimental fragmentation spectrum of the
CC   unknown peptide with a theoretical fragmentation spectra calculated for
CC   a peptide library composed of all possible linear sequences of amino
CC   acids having a total mass that corresponds to the molecular mass of the
CC   unknown peptide; and (c) identifying a peptide in the library with a
CC   theoretical fragmentation spectrum that most closely matches the
CC   fragmentation spectrum of the unknown peptide. The method is useful in
CC   DNA cloning, anti-body production, identification of recombinant
CC   products, and the study of post-translational modifications. It allows
CC   the sequence of unknown peptides or proteins with no sub-sequence
CC   identity, to be characterised using mass spectrometry. Sequences W87101
CC   to W87364 represent a library of linear peptides constructed to exemplify
CC   the method. The isoleucine residue in these peptides can be replaced by
CC   leucine to construct another 264 linear peptides to be included in the
CC   library.
XX
SQ   Sequence   6 AA;

  Query Match             37.7%;  Score 20;  DB 19;  Length 6;
  Best Local Similarity   75.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLHD 4
          | ||
Db      2 yahd 5
```

```
RESULT  34
R76987
ID   R76987 standard; peptide; 7 AA.
XX
AC   R76987;
XX
DT   09-FEB-1996  (first entry)
XX
DE   Immune stimulating peptide #4.
XX
KW   Stimulation; immune system; soybean; protease; antibody; anaphylaxis;
KW   lox toxicity.
XX
OS   Glycine max.
XX
PN   JP07138287-A.
XX
PD   30-MAY-1995.
XX
PF   02-OCT-1991;   91JP-0323606.
XX
PR   02-OCT-1991;   91JP-0323606.
XX
PA   (UCHI/) UCHIYAMA F.
XX
DR   WPI; 1995-228693/30.
XX
PT   Novel peptide immune stimulating agents - do not cause antibody
PT   production or anaphylaxis.
XX
PS   Claim 7; Page 2; 12pp; Japanese.
XX
CC   The sequences given in R76984-89 are peptides which act to stimulate
CC   the immune system.  These peptides may be prepared by chemical
CC   methods or may be isolated from soybean protease.  These peptides do
CC   not cause antibody production or anaphylaxis.  The peptides are
CC   broken down by host proteases and have, therefore, lox toxicity and
CC   are extremely safe.
XX
SQ   Sequence   7 AA;

  Query Match             37.7%;  Score 20;  DB 16;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 DPEFNL 9
          |  |||
Db      2 dkpfnl 7


RESULT  35
W07489
ID   W07489 standard; peptide; 7 AA.
XX
AC   W07489;
XX
DT   27-JAN-1997  (first entry)
XX
DE   Soybean pepsin digest derived immune activating peptide (4).
XX
KW   Soybean; immune activator; low toxicity; pepsin digest.
XX
OS   Glycine max.
XX
PN   JP08225594-A.
XX
PD   03-SEP-1996.
XX
PF   21-FEB-1995;   95JP-0072224.
XX
PR   21-FEB-1995;   95JP-0072224.
```

```
XX
PA   (SUET/) SUETSUNA Y.
XX
DR   WPI; 1996-450961/45.
XX
PT   New peptide(s) isolated from pepsin digest of soybean - have immune
PT   activating activity and low toxicity
XX
PS   Claim 7; Page 2; 16pp; Japanese.
XX
CC   The present peptide is a soybean derived peptide, with immune
CC   activating activity and low toxicity.
CC   It was prepd. by adding pepsin to a homogenate comprising 1 litre
CC   of deionised water and 200 g of soybean, and ultrafiltering the
CC   resultant hydrolysate. The filtrate was then conc. in vacuo and
CC   subjected to gel filtration by DOWEX 50WX4 and Sephadex G-25. The
CC   active fraction was then subjected to HPLC to give 8 active
CC   fractions, one of which being the present peptide.
CC   The peptide may also be prepd. by chemical synthesis using solid
CC   phase methods.
XX
SQ   Sequence   7 AA;


  Query Match             37.7%;  Score 20;  DB 17;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        4 DPEFNL 9
            |  |||
Db        2 dkpfnl 7


Search completed: December 16, 2000, 03:07:21
Job time: 8062 sec
```

```
                    GenCore version 4.5
           Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 01:51:17 ; Search time 91.14 Seconds
                                               (without alignments)
                                               6.267 Million cell updates/sec

Title:           US-09-529-121-9
Perfect score:   46
Sequence:        1 YLNVQDLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters:      182106

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_65:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                   %
Result           Query
  No.   Score   Match Length DB  ID                  Description
-----------------------------------------------------------------------
    1      46   100.0    240  2  JQ2161              dTMP kinase (EC 2.
    2      36    78.3    170  2  T02707              SKP1 protein-like
    3      36    78.3    314  2  T45077              ornithine carbamoy
    4      36    78.3    317  2  G71119              probable ornithine
    5      36    78.3    317  2  G75041              ornithine carbamoy
    6      36    78.3   1236  2  T18459              hypothetical prote
    7      35    76.1    373  2  F81438              probable periplasm
    8      35    76.1   3351  2  T13812              lipophorin - fruit
    9      34    73.9    421  2  T31787              hypothetical prote
   10      34    73.9    440  2  I39847              cell-division prot
   11      34    73.9    669  2  S64795              suppressor protein
   12      34    73.9    931  2  S66574              transferrin-bindin
   13      33    71.7    442  2  T18507              hypothetical prote
   14      33    71.7    469  2  T48314              hypothetical prote
   15      33    71.7    612  2  E81287              probable sugar tra
   16      33    71.7    710  2  T49516              Atu related protei
   17      33    71.7    981  2  S55132              hypothetical prote
   18      32    69.6    308  2  T48525              hypothetical prote
   19      32    69.6    385  2  G72561              hypothetical prote
   20      32    69.6    530  2  T10628              hypothetical prote
   21      32    69.6    532  2  T10624              reticuline oxidase
   22      32    69.6    570  2  S07744              NADH dehydrogenase
   23      32    69.6    608  2  T22897              hypothetical prote
   24      32    69.6    717  2  T22938              hypothetical prote
   25      32    69.6    942  2  T02446              hypothetical prote
   26      32    69.6   1118  2  T27865              hypothetical prote
   27      32    69.6   1436  2  B70520              probable PPE prote
   28      32    69.6   2958  2  S64921              probable membrane
   29      31    67.4    154  2  T47825              Skp1-like protein
```

```
   30      31    67.4    155  1  F64145              hypothetical prote
   31      31    67.4    179  1  G64069              uracil phosphoribo
   32      31    67.4    275  2  B40630              orf275 - Escherich
   33      31    67.4    354  2  T04779              hypothetical prote
   34      31    67.4    363  2  C69828              conserved hypothet
   35      31    67.4    375  2  T03325              gene 118 protein -
   36      31    67.4    385  2  F70448              conserved hypothet
   37      31    67.4    409  3  T40633              hypothetical prote
   38      31    67.4    456  2  T29948              hypothetical prote
   39      31    67.4    532  2  A35149              ipaH protein - Shi
   40      31    67.4    567  2  S63452              hypothetical prote
   41      31    67.4    571  2  T43288              probable exodeoxyr
   42      31    67.4    611  2  S09142              ND5 intron 1 prote
   43      31    67.4    877  2  S65057              alpha-glucosidase
   44      31    67.4   1000  2  T21970              hypothetical prote
   45      31    67.4   1374  2  S62524              ribonuclease SPAC8

                              ALIGNMENTS


RESULT   1
JQ2161
dTMP kinase (EC 2.7.4.9) - African swine fever virus
N;Alternate names: thymidylate kinase
C;Species: African swine fever virus, ASFV
C;Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 18-Jun-1999
C;Accession: JQ2161
R;Yanez, R.J.; Rodriguez, J.M.; Rodriguez, J.F.; Salas, M.L.; Vinuela, E.
J. Gen. Virol. 74, 1633-1638, 1993
A;Title: African swine fever virus thymidylate kinase gene; sequence and transcriptio
A;Reference number: JQ2161; MUID:93346971
A;Accession: JQ2161
A;Molecule type: DNA
A;Residues: 1-240 <YAN>
A;Cross-references: EMBL:Z19544; NID:g402549; PIDN:CAA79604.1; PID:g402550
A;Experimental source: strain BA71V
C;Comment: This enzyme catalyzes the phosphorylation of dTMP to dTDP.
C;Genetics:
A;Gene: A240L
C;Superfamily: dTMP kinase
C;Keywords: ATP; nucleotide biosynthesis; P-loop; phosphotransferase
F;10-17/Region: nucleotide-binding motif A (P-loop)
F;89-94/Region: nucleotide-binding motif B #status atypical


  Query Match             100.0%;  Score 46;  DB 2;  Length 240;
  Best Local Similarity   100.0%;  Pred. No. 0.052;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLNL 9
             |||||||||
Db     206 YLNVQDLNL 214


RESULT   2
T02707
SKP1 protein-like protein T18E12.14 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 23-Jul-1999
C;Accession: T02707
R;Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;
submitted to the EMBL Data Library, September 1998
A;Description: Arabidopsis thaliana chromosome II BAC T18E12 genomic sequence.
A;Reference number: Z14702
A;Accession: T02707
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-170 <ROU>
A;Cross-references: EMBL:AC005313; NID:g3548797; PID:g3548811
A;Experimental source: cultivar Columbia
C;Genetics:
A;Map position: 2
```

```
A;Note: T18E12.14


  Query Match             78.3%;  Score 36;  DB 2;  Length 170;
  Best Local Similarity   100.0%;  Pred. No. 4;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLNVQDL 7
           |||||||
Db     114 YLNVQDL 120


RESULT   3
T45077
ornithine carbamoyltransferase [imported] - Pyrococcus furiosus
C;Species: Pyrococcus furiosus
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 04-Mar-2000
C;Accession: T45077
R;Roovers, M.; Hetcke, C.; Thomm, M.; Glansdorff, N.
submitted to the EMBL Data Library, July 1996
A;Description: Enzymatic carbamoylation at high temperatures. Isolation of the gene enc
o.
A;Reference number: Z22907
A;Accession: T45077
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-314 <ROO>
A;Cross-references: EMBL:X99225; PIDN:CAA67609.1
C;Genetics:
A;Gene: argF
C;Superfamily: ornithine carbamoyltransferase; aspartate/ornithine carbamoyltransferase


  Query Match             78.3%;  Score 36;  DB 2;  Length 314;
  Best Local Similarity   77.8%;  Pred. No. 8;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLNL 9
           ||| ||| |
Db      77 YLNAQDLQL 85


RESULT   4
G71119
probable ornithine carbamoyltransferase - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C;Accession: G71119
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137
A;Accession: G71119
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-317 <KAW>
A;Cross-references: GB:AP000003; NID:g3236130; PIDN:BAA29817.1; PID:g3257134
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH0726
C;Superfamily: ornithine carbamoyltransferase; aspartate/ornithine carbamoyltransferase
F;8-310/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>


  Query Match             78.3%;  Score 36;  DB 2;  Length 317;
  Best Local Similarity   77.8%;  Pred. No. 8.1;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLNL 9
           ||| ||| |
Db      77 YLNAQDLQL 85
```

```
RESULT   5
G75041
ornithine carbamoyltransferase (argf) PAB1502 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: G75041
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A;Reference number: A75001
A;Accession: G75041
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-317 <KAW>
A;Cross-references: GB:AJ248287; GB:AL096836; NID:g5458657; PIDN:CAB50228.1; PID:g545
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB1502
C;Superfamily: ornithine carbamoyltransferase; aspartate/ornithine carbamoyltransfera
F;8-310/Domain: aspartate/ornithine carbamoyltransferase homology <AOC>


  Query Match             78.3%;  Score 36;  DB 2;  Length 317;
  Best Local Similarity   77.8%;  Pred. No. 8.1;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLNL 9
           ||| ||| |
Db      77 YLNAQDLQL 85


RESULT   6
T18459
hypothetical protein C0515c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18459
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, November 1998
A;Reference number: Z18937
A;Accession: T18459
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1236 <LAW>
A;Cross-references: EMBL:AL008970; NID:e1407852; PID:e1332552; PIDN:CAA15601.1
C;Genetics:
A;Map position: 3
A;Note: C0515c


  Query Match             78.3%;  Score 36;  DB 2;  Length 1236;
  Best Local Similarity   75.0%;  Pred. No. 39;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLN 8
           |||||::|
Db     695 YLNVQEIN 702


RESULT   7
F81438
probable periplasmic protein Cj0199c [imported] - Campylobacter jejuni (strain NCTC 1
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C;Accession: F81438
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A;Reference number: A81250; MUID:20150912
A;Accession: F81438
```

```
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-373 <PAR>
A;Cross-references: GB:AL139074; GB:AL111168; NID:g6967505; PIDN:CAB72682.1; PID:g696769
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: Cj0199c


  Query Match             76.1%;  Score 35;  DB 2;  Length 373;
  Best Local Similarity   55.6%;  Pred. No. 16;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLNVQDLNL 9
              :||::|:||
Db        121 FLNIKDINL 129


RESULT   8
T13812
lipophorin - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C;Accession: T13812
R;Kutty, R.K.; Kutty, G.; Kambadur, R.; Duncan, T.; Koonin, E.V.; Rodriguez, I.R.; Odenw
J. Biol. Chem. 271, 20641-20649, 1996
A;Title: Molecular characterization and developmental expression of a retinoid- and fatt
A;Reference number: Z17776; MUID:96355402
A;Accession: T13812
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3351 <KUT>
A;Cross-references: EMBL:U62892; NID:g1511645; PID:g1511646; PIDN:AAC47284.1
C;Genetics:
A;Map position: 4


  Query Match             76.1%;  Score 35;  DB 2;  Length 3351;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 NVQDLNL 9
              |||||||
Db        726 NVQDLNL 732


RESULT   9
T31787
hypothetical protein F44C8.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T31787
R;Dante, M.; Kramer, J.
submitted to the EMBL Data Library, July 1997
A;Description: The sequence of C. elegans cosmid F44C8.
A;Reference number: Z21086
A;Accession: T31787
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-421 <DAN>
A;Cross-references: EMBL:AF016438; PIDN:AAB65896.1; GSPDB:GN00023; CESP:F44C8.9
A;Experimental source: strain Bristol N2; clone F44C8
C;Genetics:
A;Gene: CESP:F44C8.9
A;Map position: 5
A;Introns: 61/2; 81/1; 173/2; 235/3; 299/2


  Query Match             73.9%;  Score 34;  DB 2;  Length 421;
  Best Local Similarity   75.0%;  Pred. No. 29;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LNVQDLNL 9
```

```
              |||||:|:
Db        277 LNVQDINV 284


RESULT  10
I39847
cell-division protein (septum formation) ftsA - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 20-Jun-2000
C;Accession: I39847; C69627; I40403
R;Beall, B.; Lowe, M.; Lutkenhaus, J.
J. Bacteriol. 170, 4855-4864, 1988
A;Title: Cloning and characterization of Bacillus subtilis homologs of Escherichia co
A;Reference number: I39846; MUID:89008108
A;Accession: I39847
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-440 <RES>
A;Cross-references: GB:M22630; NID:g142938; PIDN:AAA22456.1; PID:g142940
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033
A;Accession: C69627
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-440 <KUN>
A;Cross-references: GB:Z99111; GB:AL009126; NID:g2633699; PIDN:CAB13401.1; PID:g26338
A;Experimental source: strain 168
R;Gonzy-Treboul, G.; Karmazyn-Campelli, C.; Stragier, P.
J. Mol. Biol. 224, 967-979, 1992
A;Title: Developmental regulation of transcription of the Bacillus subtilis ftsAZ ope
A;Reference number: I40402; MUID:92235865
A;Accession: I40403
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-11,'I',13-14 <RE2>
A;Cross-references: EMBL:X66239; NID:g396503; PIDN:CAA46968.1; PID:g396505
C;Genetics:
A;Gene: ftsA
C;Superfamily: cell division protein ftsA


  Query Match             73.9%;  Score 34;  DB 2;  Length 440;
  Best Local Similarity   62.5%;  Pred. No. 31;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLNVQDLN 8
              |:|:|| |
Db         83 YINIQDTN 90


RESULT  11
S64795
suppressor protein FPS1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein L0729; protein YLL043w
C;Species: Saccharomyces cerevisiae
C;Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 06-Feb-1998
C;Accession: S64795; S16712
R;Wedler, H.; Wedler, E.; Scharfe, M.; Wambutt, R.
submitted to the Protein Sequence Database, May 1996
```

```
A;Reference number: S64792
A;Accession: S64795
A;Molecule type: DNA
A;Residues: 1-669 <WED>
A;Cross-references: EMBL:Z73148; NID:g1360238; PID:e245756; PID:g1360239; MIPS:YLL043w
A;Experimental source: strain S288C
R;van Aelst, L.; Hohmann, S.; Zimmermann, F.K.; Jans, A.W.H.; Thevelein, J.M.
EMBO J. 10, 2095-2104, 1991
A;Title: A yeast homologue of the bovine lens fibre MIP gene family complements the grow
mediated cAMP signalling.
A;Reference number: S16712; MUID:91293082
A;Accession: S16712
A;Molecule type: DNA
A;Residues: 1-639,'R',641-669 <AEL>
A;Cross-references: EMBL:X54157; NID:g3703; PID:g3704
C;Genetics:
A;Gene: SGD:FPS1
A;Cross-references: SGD:S0003966; MIPS:YLL043w
A;Map position: 12L
C;Keywords: transmembrane protein
F;264-280/Domain: transmembrane #status predicted <TM1>
F;326-342/Domain: transmembrane #status predicted <TM2>
F;374-390/Domain: transmembrane #status predicted <TM3>
F;455-471/Domain: transmembrane #status predicted <TM4>
F;510-526/Domain: transmembrane #status predicted <TM5>


  Query Match             73.9%;  Score 34;  DB 2;  Length 669;
  Best Local Similarity   62.5%;  Pred. No. 49;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLN 8
             |:| ||:|
Db     119 YINTQDIN 126


RESULT  12
S66574
transferrin-binding protein 1 precursor - Actinobacillus pleuropneumoniae
C;Species: Actinobacillus pleuropneumoniae
C;Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 26-Aug-1999
C;Accession: S66574; S55393
R;Daban, M.; Medrano, A.; Querol, E.
Biochem. J. 315, 257-264, 1996
A;Title: Cloning, sequencing and expression of the transferrin-binding protein 1 gene fr
A;Reference number: S66574; MUID:96207589
A;Accession: S66574
A;Molecule type: DNA
A;Residues: 1-931 <DAB>
A;Cross-references: EMBL:Z49708; NID:g853708; PIDN:CAA89810.1; PID:g853709
C;Genetics:
A;Gene: tbp1
C;Superfamily: bacterial pathogen transferrin-binding protein; tonB-dependent receptor a
C;Keywords: iron transport; membrane protein; metal binding; receptor
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-931/Product: transferrin-binding protein 1 #status predicted <MAT>
F;67-203/Domain: tonB-dependent receptor amino-terminal homology <TNN>
F;600-931/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>


  Query Match             73.9%;  Score 34;  DB 2;  Length 931;
  Best Local Similarity   66.7%;  Pred. No. 72;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLNL 9
             | ||||: |
Db     742 YFNVQDIKL 750


RESULT  13
T18507
hypothetical protein C0720w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18507
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18507
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-442 <LAW>
A;Cross-references: EMBL:Z98551; NID:e1331903; PID:e1331928; PIDN:CAB11146.1
C;Genetics:
A;Map position: 3
A;Note: C0720w


  Query Match             71.7%;  Score 33;  DB 2;  Length 442;
  Best Local Similarity   66.7%;  Pred. No. 49;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLNL 9
             | || |||:
Db     374 YQNVDDLNI 382


RESULT  14
T48314
hypothetical protein F9G14.240 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T48314
R;Bevan, M.; Terryn, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24491
A;Accession: T48314
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-469 <BEV>
A;Cross-references: EMBL:AL162973
A;Experimental source: cultivar Columbia; BAC clone F9G14
C;Genetics:
A;Map position: 5
A;Introns: 315/3; 374/3
A;Note: F9G14.240


  Query Match             71.7%;  Score 33;  DB 2;  Length 469;
  Best Local Similarity   75.0%;  Pred. No. 53;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLN 8
             ||: ||||
Db     381 YLDAQDLN 388


RESULT  15
E81287
probable sugar transferase Cj1421c [imported] - Campylobacter jejuni (strain NCTC 111
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C;Accession: E81287
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A;Reference number: A81250; MUID:20150912
A;Accession: E81287
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-612 <PAR>
A;Cross-references: GB:AL139078; GB:AL111168; NID:g6968723; PIDN:CAB73845.1; PID:g696
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
```

```
A;Gene: Cj1421c

  Query Match             71.7%;  Score 33;  DB 2;  Length 612;
  Best Local Similarity   75.0%;  Pred. No. 72;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLNVQDLN 8
          ||| :|||
Db    459 YLNKEDLN 466


Search completed: December 16, 2000, 01:51:19
Job time: 7640 sec
```

```
                    GenCore version 4.5
           Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 03:05:21 ; Search time 69.45 Seconds
                                              (without alignments)
                                              4.140 Million cell updates/sec

Title:          US-09-529-121-9
Perfect score:  46
Sequence:       1 YLNVQDLNL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters:      87993

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_39:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                   %
Result            Query
  No.    Score    Match Length DB  ID                       Description
-----------------------------------------------------------------------------
    1       46    100.0    240  1  KTHY_ASFB7               P42490 african swi
    2       36     78.3    315  1  OTC_PYRFU                Q51742 pyrococcus
    3       36     78.3    317  1  OTC_PYRAB                O93656 pyrococcus
    4       36     78.3    317  1  OTC_PYRHO                O58457 pyrococcus
    5       34     73.9    440  1  FTSA_BACSU               P28264 bacillus su
    6       34     73.9    669  1  FPS1_YEAST               P23900 saccharomyc
    7       33     71.7    981  1  YM48_YEAST               Q12751 saccharomyc
    8       32     69.6    570  1  NU5M_PARTE               P15584 paramecium
    9       31     67.4    155  1  YHCH_HAEIN               P44583 haemophilus
   10       31     67.4    179  1  Y459_HAEIN               P44722 haemophilus
   11       31     67.4    398  1  HISX_SULSO               O33775 sulfolobus
   12       31     67.4    455  1  PEX3_PICPA               Q92262 pichia past
   13       31     67.4    532  1  IPA7_SHIFL               P18014 shigella fl
   14       31     67.4    571  1  EXO1_SCHPO               P53695 schizosacch
   15       31     67.4    732  1  KMHB_DICDI               P90648 dictyosteli
   16       31     67.4    877  1  AGLU_HORVU               Q43763 hordeum vul
   17       31     67.4   1000  1  COPP_CAEEL               Q20168 caenorhabdi
   18       31     67.4   1374  1  YC9A_SCHPO               Q09884 schizosacch
   19       31     67.4   2109  1  RRPL_VSVJO               P16379 vesicular s
   20       30     65.2    147  1  Y978_METJA               Q58388 methanococc
   21       30     65.2    185  1  DUS3_HUMAN               P51452 homo sapien
   22       30     65.2    247  1  Y125_RICPR               P41087 rickettsia
   23       30     65.2    367  1  Y669_METJA               Q58083 methanococc
   24       30     65.2    390  1  ACKA_MYCPN               P75245 mycoplasma
   25       30     65.2    486  1  YAIT_ECOLI               P77199 escherichia
   26       30     65.2    557  1  APG6_YEAST               Q02948 saccharomyc
   27       30     65.2    589  1  TRPG_THEMA               Q08654 thermotoga
   28       30     65.2    623  1  VGLG_BEFV                P32595 bovine ephe
   29       30     65.2    626  1  NR43_HUMAN               Q92570 homo sapien
   30       30     65.2    628  1  NR43_RAT                 P51179 rattus norv
   31       30     65.2    719  1  MUS2_STAAU               Q9zeh5 staphylococ
   32       30     65.2    750  1  MDOB_ECOLI               P39401 escherichia
   33       30     65.2    897  1  BGAL_CLOAB               P24131 clostridium
   34       30     65.2   1246  1  VP03_HSVSA               Q01000 herpesvirus
   35       30     65.2   1256  1  ATL_STAAU                P52081 staphylococ
   36       30     65.2   2452  1  RPB1_PLAFD               P14248 plasmodium
   37       30     65.2   3063  1  CA1C_HUMAN               Q99715 homo sapien
   38       30     65.2   3067  1  CA1C_MOUSE               Q60847 mus musculu
   39       30     65.2   3124  1  CA1C_CHICK               P13944 gallus gall
   40       30     65.2   5147  1  FAT_DROME                P33450 drosophila
   41     29.5     64.1    460  1  YEGQ_HAEIN               P44700 haemophilus
   42       29     63.0    160  1  YS57_CAEEL               Q09665 caenorhabdi
   43       29     63.0    197  1  BINL_STAAU               P18358 staphylococ
   44       29     63.0    211  1  GDNF_HUMAN               P39905 homo sapien
   45       29     63.0    211  1  GDNF_MOUSE               P48540 mus musculu

                                   ALIGNMENTS

RESULT   1
KTHY_ASFB7
ID   KTHY_ASFB7     STANDARD;      PRT;   240 AA.
AC   P42490;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   THYMIDYLATE KINASE (EC 2.7.4.9) (DTMP KINASE).
GN   A240L.
OS   African swine fever virus (strain BA71V) (ASFV).
OC   Viruses; dsDNA viruses, no RNA stage;
OC   African swine fever-like viruses.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93346971.
RA   Yanez R.J., Rodriguez J.M., Rodriguez J.F., Salas M.L., Vinuela E.;
RT   "African swine fever virus thymidylate kinase gene: sequence and
RT   transcriptional mapping.";
RL   J. Gen. Virol. 74:1633-1638(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,
RA   Rodriguez J.F., Vinuela E.;
RT   "Analysis of the complete nucleotide sequence of African swine fever
RT   virus.";
RL   Virology 208:249-278(1995).
CC   -!- CATALYTIC ACTIVITY: ATP + THYMIDINE 5'-MONOPHOSPHATE = ADP +
CC       THYMIDINE 5'-DIPHOSPHATE.
CC   -!- PATHWAY: SYNTHESIS OF DTTP FROM DTMP.
CC   -!- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z19544; CAA79604.1; -.
DR   EMBL; U18466; AAA65265.1; -.
DR   HSSP; P00572; 2TMK.
DR   INTERPRO; IPR000062; -.
DR   PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
KW   Transferase; Kinase; Nucleotide biosynthesis; ATP-binding.
FT   NP_BIND      10     17       ATP (POTENTIAL).
SQ   SEQUENCE   240 AA;  27755 MW;  AF89DBE6CCE3E44D CRC64;


  Query Match             100.0%;  Score 46;  DB 1;  Length 240;
  Best Local Similarity   100.0%;  Pred. No. 0.022;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLNL 9
             |||||||||
Db     206 YLNVQDLNL 214
```

```
RESULT   2
OTC_PYRFU
ID   OTC_PYRFU      STANDARD;      PRT;   315 AA.
AC   Q51742;
DT   01-NOV-1997 (Rel. 35, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ORNITHINE CARBAMOYLTRANSFERASE (EC 2.1.3.3) (OTCASE).
GN   ARGF.
OS   Pyrococcus furiosus.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 3638 / VC1;
RX   MEDLINE; 97433289.
RA   Roovers M., Hetcke C., Legrain C., Thomm M., Glansdorff N.;
RT   "Isolation of the gene encoding Pyrococcus furiosus ornithine
RT   carbamoyltransferase and study of its expression profile in vivo and
RT   in vitro.";
RL   Eur. J. Biochem. 247:1038-1045(1997).
RN   [2]
RP   REVISION TO 207.
RA   Roovers M.;
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RX   MEDLINE; 98169451.
RA   Villeret V., Clantin B., Tricot C., Legrain C., Roovers M., Stalon V.,
RA   Glansdorff N., van Beeumen J.;
RT   "The crystal structure of Pyrococcus furiosus ornithine
RT   carbamoyltransferase reveals a key role for oligomerization in enzyme
RT   stability at extremely high temperatures.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:2801-2806(1998).
CC   -!- CATALYTIC ACTIVITY: CARBAMOYL PHOSPHATE + ORNITHINE = CITRULLINE
CC       + ORTHOPHOSPHATE.
CC   -!- PATHWAY: ARGININE BIOSYNTHESIS (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X99225; CAA67609.1; -.
DR   EMBL; Y12727; CAA73260.1; -.
DR   PDB; 1A1S; 17-JUN-98.
DR   INTERPRO; IPR002029; -.
DR   INTERPRO; IPR002082; -.
DR   INTERPRO; IPR002292; -.
DR   PFAM; PF00185; OTCace; 1.
DR   PRINTS; PR00100; AOTCASE.
DR   PRINTS; PR00101; ATCASE.
DR   PRINTS; PR00102; OTCASE.
DR   PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
KW   Transferase; Arginine biosynthesis; 3D-structure.
FT   CONFLICT    207    207       MISSING (IN CAA67609).
SQ   SEQUENCE   315 AA;  35280 MW;  62804E71F2A88E81 CRC64;

  Query Match             78.3%;  Score 36;  DB 1;  Length 315;
  Best Local Similarity   77.8%;  Pred. No. 3.5;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 YLNVQDLNL 9
          ||| ||| |
Db     77 YLNAQDLQL 85
```

```
RESULT   3
OTC_PYRAB
ID   OTC_PYRAB      STANDARD;      PRT;   317 AA.
AC   O93656;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ORNITHINE CARBAMOYLTRANSFERASE (EC 2.1.3.3) (OTCASE).
GN   ARGF OR PAB1502.
OS   Pyrococcus abyssi.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GE5;
RA   Cunin R., Seumois G., Purcarea C., Van Vliet F., Legrain C.;
RT   "Ornithine carbamoyltransferase from Pyrococcus abyssi.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ORSAY;
RA   Heilig R.;
RT   "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT   structure and evolution.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: CARBAMOYL PHOSPHATE + ORNITHINE = CITRULLINE
CC       + ORTHOPHOSPHATE.
CC   -!- PATHWAY: ARGININE BIOSYNTHESIS (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF083209; AAD09004.1; -.
DR   EMBL; AJ248287; CAB50228.1; -.
DR   HSSP; Q51742; 1A1S.
DR   INTERPRO; IPR002029; -.
DR   INTERPRO; IPR002082; -.
DR   INTERPRO; IPR002292; -.
DR   PFAM; PF00185; OTCace; 1.
DR   PRINTS; PR00100; AOTCASE.
DR   PRINTS; PR00101; ATCASE.
DR   PRINTS; PR00102; OTCASE.
DR   PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
KW   Transferase; Arginine biosynthesis.
FT   CONFLICT     63     63       S -> R (IN REF. 1).
FT   CONFLICT     75     75       A -> G (IN REF. 1).
FT   CONFLICT    159    159       V -> I (IN REF. 1).
FT   CONFLICT    201    202       EQ -> DE (IN REF. 1).
FT   CONFLICT    264    264       D -> S (IN REF. 1).
FT   CONFLICT    313    317       VKTGF -> RKDGLLTF (IN REF. 1).
SQ   SEQUENCE   317 AA;  35357 MW;  F4686B091F76494F CRC64;

  Query Match             78.3%;  Score 36;  DB 1;  Length 317;
  Best Local Similarity   77.8%;  Pred. No. 3.5;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 YLNVQDLNL 9
          ||| ||| |
Db     77 YLNAQDLQL 85

RESULT   4
OTC_PYRHO
ID   OTC_PYRHO      STANDARD;      PRT;   317 AA.
AC   O58457;
```

```
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ORNITHINE CARBAMOYLTRANSFERASE (EC 2.1.3.3) (OTCASE).
GN   ARGF OR PH0726.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE; 98344137.
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
CC   -!- CATALYTIC ACTIVITY: CARBAMOYL PHOSPHATE + ORNITHINE = CITRULLINE
CC       + ORTHOPHOSPHATE.
CC   -!- PATHWAY: ARGININE BIOSYNTHESIS (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AP000003; BAA29817.1; -.
DR   INTERPRO; IPR002029; -.
DR   INTERPRO; IPR002082; -.
DR   INTERPRO; IPR002292; -.
DR   PFAM; PF00185; OTCace; 1.
DR   PRINTS; PR00100; AOTCASE.
DR   PRINTS; PR00101; ATCASE.
DR   PRINTS; PR00102; OTCASE.
DR   PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
KW   Transferase; Arginine biosynthesis.
SQ   SEQUENCE   317 AA;  35368 MW;  D949989E0E8F6B1E CRC64;


  Query Match             78.3%;  Score 36;  DB 1;  Length 317;
  Best Local Similarity   77.8%;  Pred. No. 3.5;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLNL 9
           ||| ||| |
Db      77 YLNAQDLQL 85


RESULT   5
FTSA_BACSU
ID   FTSA_BACSU     STANDARD;      PRT;   440 AA.
AC   P28264; Q45573;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   CELL DIVISION PROTEIN FTSA.
GN   FTSA.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89008108.
RA   Beall B., Lowe M., Lutkenhaus J.;
RT   "Cloning and characterization of Bacillus subtilis homologs of
RT   Escherichia coli cell division genes ftsZ and ftsA.";
RL   J. Bacteriol. 170:4855-4864(1988).
RN   [2]
RP   SEQUENCE OF 1-14 FROM N.A.
RC   STRAIN=168 / JH642;
RX   MEDLINE; 92235865.
RA   Gonzy-Treboul G., Karmazyn-Campelli C., Stragier P.;
RT   "Developmental regulation of transcription of the Bacillus subtilis
RT   ftsAZ operon.";
RL   J. Mol. Biol. 224:967-979(1992).
RN   [3]
RP   SEQUENCE OF 1-8 FROM N.A.
RX   MEDLINE; 92325056.
RA   Gholamhoseinian A., Shen Z., Wu J.J., Piggot P.;
RT   "Regulation of transcription of the cell division gene ftsA during
RT   sporulation of Bacillus subtilis.";
RL   J. Bacteriol. 174:4647-4656(1992).
CC   -!- FUNCTION: THIS PROTEIN MAY BE INVOLVED IN ANOMALOUS FILAMENT
CC       GROWTH (BY SIMILARITY). MAY BE A COMPONENT OF THE SEPTUM. IT
CC       MAY INTERACT WITH FTSZ (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE FTSA/MREB FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M22630; AAA22456.1; -.
DR   EMBL; X66239; CAA46968.1; -.
DR   EMBL; S39431; AAD13818.1; -.
DR   EMBL; Z99111; CAB13401.1; -.
DR   SUBTILIST; BG10231; FTSA.
KW   Cell division; Cell shape.
FT   CONFLICT     12     12       L -> I (IN REF. 2).
SQ   SEQUENCE   440 AA;  48102 MW;  6064D55961358C44 CRC64;


  Query Match             73.9%;  Score 34;  DB 1;  Length 440;
  Best Local Similarity   62.5%;  Pred. No. 13;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLN 8
           |:|:|| |
Db      83 YINIQDTN 90


RESULT   6
FPS1_YEAST
ID   FPS1_YEAST     STANDARD;      PRT;   669 AA.
AC   P23900;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   GLYCEROL UPTAKE/EFFLUX FACILITATOR PROTEIN.
GN   FPS1 OR YLL043W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91293082.
RA   van Aelst L., Hohmann S., Zimmermann F.K., Jans A.W.H.,
RA   Thevelein J.M.;
RT   "A yeast homologue of the bovine lens fibre MIP gene family
RT   complements the growth defect of a Saccharomyces cerevisiae mutant on
RT   fermentable sugars but not its defect in glucose-induced RAS-mediated
RT   cAMP signalling.";
RL   EMBO J. 10:2095-2104(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RA   Wedler H., Wedler E., Scharfe M., Wambutt R.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE; 95246729.
RA   Luyten K., Albertyn J., Skibbe W.F., Prior B.A., Ramos J.,
RA   Thevelein J.M., Hohmann S.;
RT   "Fps1, a yeast member of the MIP family of channel proteins, is a
RT   facilitator for glycerol uptake and efflux and is inactive under
RT   osmotic stress.";
RL   EMBO J. 14:1360-1371(1995).
CC   -!- FUNCTION: CHANNEL PROTEIN FOR GLYCEROL. HAS A ROLE IN BOTH
CC       GLYCEROL INFLUX AND EFFLUX. PLAYS A ROLE IN OSMOREGULATION: UNDER
CC       OSMOTIC STREE THE CHANNEL IS APPARENTLY CLOSED TO ALLOW
CC       ACCUMULATION OF GLYCEROL IN THE CELL UNDER HYPEROSMOTIC
CC       CONDITIONS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONTAINS 6
CC       POTENTIAL TRANSMEMBRANE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE TRANSMEMBRANE CHANNEL MIP FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X54157; CAA38096.1; -.
DR   EMBL; Z73148; CAA97494.1; -.
DR   PIR; S16712; S16712.
DR   SGD; S0003966; FPS1.
DR   INTERPRO; IPR000425; -.
DR   PFAM; PF00230; MIP; 1.
DR   PRINTS; PR00783; MINTRINSICP.
DR   PROSITE; PS00221; MIP; 1.
KW   Glycerol metabolism; Transport; Transmembrane.
FT   TRANSMEM    255    275       POTENTIAL.
FT   TRANSMEM    326    346       POTENTIAL.
FT   TRANSMEM    370    390       POTENTIAL.
FT   TRANSMEM    447    467       POTENTIAL.
FT   TRANSMEM    507    527       POTENTIAL.
FT   DOMAIN       50     69       ASN-RICH.
FT   CONFLICT    640    640       A -> R (IN REF. 1).
SQ   SEQUENCE   669 AA;  73877 MW;  BA9C78056A1251B9 CRC64;

  Query Match             73.9%;  Score 34;  DB 1;  Length 669;
  Best Local Similarity   62.5%;  Pred. No. 22;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLN 8
            |:| ||:|
Db      119 YINTQDIN 126


RESULT   7
YM48_YEAST
ID   YM48_YEAST     STANDARD;      PRT;   981 AA.
AC   Q12751;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 113.2 KDA PROTEIN IN SSO2-HSC82 INTERGENIC REGION.
GN   YMR185W OR YM8010.15.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
```

```
RN   [2]
RP   SEQUENCE OF 650-981 FROM N.A.
RX   MEDLINE; 95373281.
RA   Erkine A.M., Szent-Gyorgyi C., Simmons S.F., Gross D.S.;
RT   "The upstream sequences of the HSP82 and HSC82 genes of Saccharomyces
RT   cerevisiae: regulatory elements and nucleosome positioning motifs.";
RL   Yeast 11:573-580(1995).
RN   [3]
RP   SEQUENCE OF 976-981 FROM N.A.
RX   MEDLINE; 89384620.
RA   Borkovich K.A., Farrelly F.W., Finkelstein D.B., Taulien J.,
RA   Lindquist S.;
RT   "hsp82 is an essential protein that is required in higher
RT   concentrations for growth of cells at higher temperatures.";
RL   Mol. Cell. Biol. 9:3919-3930(1989).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z49808; CAA89918.1; -.
DR   EMBL; U20349; AAA87904.1; ALT_INIT.
DR   SGD; S0004797; YMR185W.
KW   Hypothetical protein.
SQ   SEQUENCE   981 AA;  113248 MW;  13F29F65C6D4E53B CRC64;

  Query Match             71.7%;  Score 33;  DB 1;  Length 981;
  Best Local Similarity   77.8%;  Pred. No. 56;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLNL 9
            ||||  |||
Db      471 YLNVLSLNL 479


RESULT   8
NU5M_PARTE
ID   NU5M_PARTE     STANDARD;      PRT;   570 AA.
AC   P15584;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).
GN   ND5 OR NDH5.
OS   Paramecium tetraurelia.
OG   Mitochondrion.
OC   Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC   Paramecium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=STOCK 51;
RX   MEDLINE; 90174913.
RA   Pritchard A.E., Seilhamer J.J., Mahalingam R., Sable C.L.,
RA   Venuti S.E., Cummings D.J.;
RT   "Nucleotide sequence of the mitochondrial genome of Paramecium.";
RL   Nucleic Acids Res. 18:173-180(1990).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X15917; CAA34053.1; -.
DR   PIR; S07744; S07744.
```

```
DR   INTERPRO; IPR001516; -.
DR   INTERPRO; IPR001750; -.
DR   PFAM; PF00361; oxidored_q1; 1.
DR   PFAM; PF00662; oxidored_q1_N; 1.
KW   Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ   SEQUENCE   570 AA;  65196 MW;  23636CFDC1B0BC4C CRC64;


  Query Match             69.6%;  Score 32;  DB 1;  Length 570;
  Best Local Similarity   75.0%;  Pred. No. 47;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLN 8
             | || |||
Db     187 YANVHDLN 194


RESULT   9
YHCH_HAEIN
ID   YHCH_HAEIN     STANDARD;      PRT;   155 AA.
AC   P44583;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   HYPOTHETICAL PROTEIN HI0227.
GN   HI0227.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- SIMILARITY: STRONG, TO E.COLI YHCH.
CC   -!- SIMILARITY: BELONGS TO THE YHCH/YIAL/YJGK FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U32709; AAC21897.1; -.
DR   TIGR; HI0227; -.
KW   Hypothetical protein.
SQ   SEQUENCE   155 AA;  17670 MW;  DAC9FEEAE69B11C1 CRC64;


  Query Match             67.4%;  Score 31;  DB 1;  Length 155;
  Best Local Similarity   75.0%;  Pred. No. 16;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLN 8
             |||  |||
Db      25 YLNTLDLN 32


RESULT  10
Y459_HAEIN
ID   Y459_HAEIN     STANDARD;      PRT;   179 AA.
AC   P44722;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL PROTEIN HI0459.
GN   HI0459.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: COULD BE A TRANSCRIPTIONAL REGULATOR WITH A LOW ACTIVITY
CC       AS A URACIL PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.9) (UPRTASE)
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: STRONG, TO BACILLUS PYRIMIDINE OPERON REGULATORY
CC       PROTEIN PYRR.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U32728; AAC22117.1; -.
DR   HSSP; P39765; 1A4X.
DR   TIGR; HI0459; -.
DR   INTERPRO; IPR000836; -.
DR   PFAM; PF00156; Pribosyltran; 1.
KW   Hypothetical protein; Transcription regulation; DNA-binding;
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE   179 AA;  20527 MW;  1E96972790C0A9E9 CRC64;


  Query Match             67.4%;  Score 31;  DB 1;  Length 179;
  Best Local Similarity   83.3%;  Pred. No. 19;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLNVQD 6
             |||:||
Db      92 YLNIQD 97


RESULT  11
HISX_SULSO
ID   HISX_SULSO     STANDARD;      PRT;   398 AA.
AC   O33775;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HISTIDINOL DEHYDROGENASE (EC 1.1.1.23) (HDH).
GN   HISD.
OS   Sulfolobus solfataricus.
OC   Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 1617 / P2;
```

```
RX   MEDLINE; 97352708.
RA   Charlebois R.L., Sensen C.W., Doolittle W.F., Brown J.R.;
RT   "Evolutionary analysis of the hisCGABdFDEHI gene cluster from the
RT   archaeon Sulfolobus solfataricus P2.";
RL   J. Bacteriol. 179:4429-4432(1997).
CC   -!- FUNCTION: THIS PROTEIN IS CONSIDERED AS A BIFUNCTIONAL ENZYME,
CC       POSSESSING TWO ACTIVE SITES, ONE AN ALCOHOL DEHYDROGENASE AND
CC       THE OTHER AN ALDEHYDE DEHYDROGENASE (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: L-HISTIDINOL + 2 NAD(+) = L-HISTIDINE +
CC       2 NADH.
CC   -!- PATHWAY: TENTH STEP IN HISTIDINE BIOSYNTHETIC PATHWAY.
CC   -!- SIMILARITY: TO OTHER PROKARYOTIC, FUNGAL AND PLANTS HDH.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license-agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U82227; AAB63023.1; -.
DR   INTERPRO; IPR001692; -.
DR   PFAM; PF00815; Histidinol_dh; 1.
DR   PRINTS; PR00083; HOLDHDRGNASE.
DR   PROSITE; PS00611; HISOL_DEHYDROGENASE; FALSE_NEG.
KW   Histidine biosynthesis; Multifunctional enzyme; Oxidoreductase; NAD.
SQ   SEQUENCE   398 AA;  43635 MW;  B989B35227ED564B CRC64;

  Query Match             67.4%;  Score 31;  DB 1;  Length 398;
  Best Local Similarity   62.5%;  Pred. No. 50;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLN 8
            | ||:|:|
Db      363 YANVRDIN 370


RESULT 12
PEX3_PICPA
ID   PEX3_PICPA     STANDARD;      PRT;   455 AA.
AC   Q92262;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PEROXISOMAL MEMBRANE PROTEIN PAS2 (PEROXIN-3).
GN   PEX3 OR PAS2.
OS   Pichia pastoris (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Pichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97115764.
RA   Subramani S.;
RT   "Protein translocation into peroxisomes.";
RL   J. Biol. Chem. 271:32483-32486(1996).
CC   -!- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z72390; CAA96530.1; -.
KW   Transmembrane; Peroxisome.
FT   DOMAIN        1     15       MATRIX (POTENTIAL).
FT   TRANSMEM     16     33       POTENTIAL.
FT   DOMAIN       34    455       CYTOPLASMIC (POTENTIAL).
```

```
SQ   SEQUENCE   455 AA;  51973 MW;  6853C5EA5C67EC34 CRC64;

  Query Match             67.4%;  Score 31;  DB 1;  Length 455;
  Best Local Similarity   66.7%;  Pred. No. 58;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLNL 9
            |||  |:||
Db      353 YLNNADINL 361


RESULT 13
IPA7_SHIFL
ID   IPA7_SHIFL     STANDARD;      PRT;   532 AA.
AC   P18014;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   60 KDA ANTIGEN.
GN   IPAH 7.8.
OS   Shigella flexneri.
OG   Plasmid 210 kb invasion pWR100.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Shigella.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M90T / SEROTYPE 5;
RX   MEDLINE; 90202708.
RA   Hartman A.B., Venkatesan M.M., Oaks E.V., Buysse J.M.;
RT   "Sequence and molecular characterization of a multicopy invasion
RT   plasmid antigen gene, ipaH, of Shigella flexneri.";
RL   J. Bacteriol. 172:1905-1915(1990).
CC   -!- SIMILARITY: TO Y.PESTIS YOPM.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M32063; AAA26526.1; -.
DR   PIR; A35149; A35149.
DR   INTERPRO; IPR001611; -.
DR   PFAM; PF00560; LRR; 2.
KW   Antigen; Plasmid; Virulence; Repeat; Multigene family.
FT   DOMAIN       36    149       6 X 14 AA APPROXIMATE TANDEM REPEATS OF
FT                                L-X(2)-L-P-X-L-P-X(2)-L-X(2)-L.
FT   REPEAT       36     49       1.
FT   REPEAT       56     69       2.
FT   REPEAT       76     89       3.
FT   REPEAT       96    109       4.
FT   REPEAT      116    129       5.
FT   REPEAT      136    149       6.
SQ   SEQUENCE   532 AA;  60955 MW;  58188D4EF50B6917 CRC64;

  Query Match             67.4%;  Score 31;  DB 1;  Length 532;
  Best Local Similarity   75.0%;  Pred. No. 70;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        2 LNVQDLNL 9
            ||: ||||
Db       29 LNLSDLNL 36


RESULT 14
EXO1_SCHPO
ID   EXO1_SCHPO     STANDARD;      PRT;   571 AA.
AC   P53695;
DT   01-OCT-1996 (Rel. 34, Created)
```

```
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   EXONUCLEASE I (EXO I).
GN   EXO1.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC   Schizosaccharomycetaceae; Schizosaccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 139-160.
RC   STRAIN=972;
RX   MEDLINE; 95158876.
RA   Szankasi P., Smith G.R.;
RT   "A role for exonuclease I from S. pombe in mutation avoidance and
RT   mismatch correction.";
RL   Science 267:1166-1169(1995).
CC   -!- FUNCTION: 5'->3' DOUBLE-STRANDED DNA EXONUCLEASE THAT COULD ACT IN
CC       A PATHWAY THAT CORRECTS MISMATCHED BASE PAIRS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- INDUCTION: BY MEIOSIS.
CC   -!- SIMILARITY: TO OTHER MEMBERS OF THE XPG/RAD2 FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L35174; AAC41648.1; -.
DR   INTERPRO; IPR001532; -.
DR   PFAM; PF00867; XPG_I; 1.
DR   PFAM; PF00752; XPG_N; 1.
DR   PRINTS; PR00853; XPGRADSUPER.
DR   PROSITE; PS00841; XPG_1; 1.
DR   PROSITE; PS00842; XPG_2; 1.
KW   DNA repair; Hydrolase; Nuclease; Endonuclease; Nuclear protein.
FT   DOMAIN        1     96       N-DOMAIN.
FT   DOMAIN      114    245       I-DOMAIN.
SQ   SEQUENCE   571 AA;  63867 MW;  7D21B3778104BA9D CRC64;

  Query Match             67.4%;  Score 31;  DB 1;  Length 571;
  Best Local Similarity   85.7%;  Pred. No. 76;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 NVQDLNL 9
             | |||||
Db     202 NAQDLNL 208


RESULT 15
KMHB_DICDI
ID   KMHB_DICDI     STANDARD;      PRT;   732 AA.
AC   P90648;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   MYOSIN HEAVY CHAIN KINASE B (EC 2.7.1.129) (MHCK B).
GN   MHKB OR MHCKB.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Dictyosteliida; Dictyostelium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX3;
RX   MEDLINE; 97277316.
RA   Clancy C.E., Mendoza M.G., Naismith T.V., Kolman M.F., Egelhoff T.T.;
RT   "Identification of a protein kinase from Dictyostelium with homology
RT   to the novel catalytic domain of myosin heavy chain kinase A.";
RL   J. Biol. Chem. 272:11812-11815(1997).
CC   -!- FUNCTION: PHOSPHORYLATES THREONINE IN THE C-TERMINAL TAIL REGION
CC       OF MYOSIN II HEAVY CHAIN. THIS PHOSPHORYLATION IS CRITICAL IN
CC       REGULATING THE ASSEMBLY AND DISASSEMBLY OF MYOSIN II FILAMENT.
CC   -!- CATALYTIC ACTIVITY: ATP + [MYOSIN HEAVY-CHAIN] = ADP +
CC       [MYOSIN HEAVY-CHAIN] PHOSPHATE.
CC   -!- DOMAIN: CONSISTS OF AN N-TERMINAL DOMAIN, A CENTRAL NONREPETITIVE
CC       CATALYTIC DOMAIN, AND A C-TERMINAL DOMAIN WITH SEVEN WD REPEATS.
CC   -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC   -!- SIMILARITY: BELONGS TO THE MHCK / EF-2 PROTEIN KINASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U90946; AAB50136.1; -.
DR   DICTYDB; DD01087; MHKB.
DR   INTERPRO; IPR001680; -.
DR   PFAM; PF00400; WD40; 7.
DR   PRINTS; PR00320; GPROTEINBRPT.
DR   PROSITE; PS00678; WD_REPEATS; 5.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat;
KW   WD repeat.
FT   NP_BIND     298    303       ATP (POTENTIAL).
FT   DOMAIN      355    381       POLY-ASN.
FT   REPEAT      458    486       WD1.
FT   REPEAT      500    528       WD2.
FT   REPEAT      540    568       WD3.
FT   REPEAT      580    608       WD4.
FT   REPEAT      620    648       WD5.
FT   REPEAT      660    688       WD6.
FT   REPEAT      700    730       WD7.
SQ   SEQUENCE   732 AA;  83167 MW;  A7233C4BD56D4088 CRC64;

  Query Match             67.4%;  Score 31;  DB 1;  Length 732;
  Best Local Similarity   62.5%;  Pred. No. 1e+02;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLN 8
             |||:| :|
Db     322 YLNLQSIN 329


Search completed: December 16, 2000, 03:05:22
Job time: 8042 sec
```

```
                        GenCore version 4.5
               Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 02:01:00 ; Search time 113.2 Seconds
                                               (without alignments)
                                               7.424 Million cell updates/sec

Title:           US-09-529-121-9
Perfect score:   46
Sequence:        1 YLNVQDLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters:      297973

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_14:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                   %
Result          Query
  No.   Score   Match Length DB  ID                              Description
 --------------------------------------------------------------------------------
    1      36    78.3     170 10  O81055                           O81055 arabidopsis
    2      36    78.3    1236 5   O77346                          O77346 plasmodium
    3      35    76.1     175 5   Q9W125                          Q9w125 drosophila
    4      35    76.1     303 2   P71441                          P71441 leptospira
    5      35    76.1     303 2   Q9S4F5                          Q9s4f5 leptospira
    6      35    76.1    2307 5   Q9V496                          Q9v496 drosophila
    7      35    76.1    3351 5   Q94907                          Q94907 drosophila
    8      34    73.9     421 5   O16355                          O16355 caenorhabdi
    9      34    73.9     898 2   Q44125                          Q44125 actinobacil
   10      34    73.9     898 2   Q44158                          Q44158 actinobacil
   11      34    73.9     931 2   Q44127                          Q44127 actinobacil
   12      34    73.9     931 2   Q44166                          Q44166 actinobacil
   13      33    71.7     442 5   O77390                          O77390 plasmodium
   14      33    71.7     442 8   Q9XMT9                          Q9xmt9 tetrahymena
   15      33    71.7     997 10  Q9SJK3                           Q9sjk3 arabidopsis
   16      33    71.7    1464 12  Q66951                           Q66951 feline infe
   17      33    71.7    1820 5   Q9VVG0                          Q9vvg0 drosophila
   18      32    69.6     362 3   O74224                          O74224 emericella
   19      32    69.6     385 1   Q9YB18                          Q9yb18 aeropyrum p
```

```
   20      32    69.6     468 11  O09040                          O09040 mus musculu
   21      32    69.6     530 10  Q9SUC6                           Q9suc6 arabidopsis
   22      32    69.6     532 10  Q9SVG5                           Q9svg5 arabidopsis
   23      32    69.6     608 5   Q20969                          Q20969 caenorhabdi
   24      32    69.6     717 5   Q9XVM1                          Q9xvm1 caenorhabdi
   25      32    69.6     748 5   Q9W1L4                          Q9w114 drosophila
   26      32    69.6     791 4   O75120                          O75120 homo sapien
   27      32    69.6     942 10  O82780                           O82780 arabidopsis
   28      32    69.6    1118 5   Q23496                          Q23496 caenorhabdi
   29      32    69.6    1436 2   O07716                          O07716 mycobacteri
   30      32    69.6    2958 3   Q12150                          Q12150 saccharomyc
   31      32    69.6    3906 2   Q9RNB2                          Q9rnb2 microcystis
   32      31    67.4     175 12  Q89031                           Q89031 vesicular s
   33      31    67.4     275 2   Q03084                          Q03084 escherichia
   34      31    67.4     301 5   Q9VXH7                          Q9vxh7 drosophila
   35      31    67.4     303 2   Q9ZGJ5                          Q9zgj5 leptospira
   36      31    67.4     308 2   O07114                          O07114 mastigoclad
   37      31    67.4     337 2   Q9RP56                          Q9rp56 escherichia
   38      31    67.4     354 10  Q9SYZ8                           Q9syz8 arabidopsis
   39      31    67.4     363 2   O07544                          O07544 bacillus su
   40      31    67.4     375 9   O80118                          O80118 bacteriopha
   41      31    67.4     385 2   O67616                          O67616 aquifex aeo
   42      31    67.4     399 13  O93458                           O93458 podarcis si
   43      31    67.4     403 5   Q9VVY5                          Q9vvy5 drosophila
   44      31    67.4     409 3   Q9Y7L4                          Q9y714 schizosacch
   45      31    67.4     456 5   P91469                          P91469 caenorhabdi

                                  ALIGNMENTS


RESULT   1
O81055
ID   O81055      PRELIMINARY;      PRT;   170 AA.
AC   O81055;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   KINETOCHORE (SKP1P)-LIKE PROTEIN.
GN   T18E12.14.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC   Brassicaceae; Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
RA   Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.;
RT   "Arabidopsis thaliana chromosome II BAC T18E12 genomic sequence.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005313; AAC34483.1; -.
DR   MENDEL; 31709; Arath;1266;31709.
DR   INTERPRO; IPR001232; -.
DR   PFAM; PF01466; Skp1; 1.
SQ   SEQUENCE   170 AA;  19398 MW;  B75BEB9552CF7048 CRC64;

  Query Match             78.3%;  Score 36;  DB 10;  Length 170;
  Best Local Similarity   100.0%;  Pred. No. 9.6;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLNVQDL 7
           |||||||
Db     114 YLNVQDL 120


RESULT   2
O77346
ID   O77346      PRELIMINARY;      PRT;  1236 AA.
AC   O77346;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
```

```
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   PFC0515C PROTEIN.
GN   PFC0515C.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RA   Hamlin N., Lawson D., Barrell B.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL008970; CAA15601.1; -.
DR   INTERPRO; IPR001440; -.
SQ   SEQUENCE   1236 AA;  147954 MW;  011A46685F526E24 CRC64;

  Query Match             78.3%;  Score 36;  DB 5;  Length 1236;
  Best Local Similarity   75.0%;  Pred. No. 81;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLN 8
           |||||::|
Db     695 YLNVQEIN 702

RESULT   3
Q9W125
ID   Q9W125      PRELIMINARY;      PRT;   175 AA.
AC   Q9W125;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG3683 PROTEIN.
GN   CG3683.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE; 20196006.
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
```

```
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003464; AAF47252.1; -.
DR   FLYBASE; FBgn0035046; CG3683.
SQ   SEQUENCE   175 AA;  19845 MW;  508171EEDDF4D9DE CRC64;

  Query Match             76.1%;  Score 35;  DB 5;  Length 175;
  Best Local Similarity   87.5%;  Pred. No. 16;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LNVQDLNL 9
             ||||:|||
Db      15 LNVQELNL 22

RESULT   4
P71441
ID   P71441      PRELIMINARY;      PRT;   303 AA.
AC   P71441;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   RFBF.
GN   RFBF.
OS   Leptospira interrogans.
OC   Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COPENHAGENI (L45);
RX   MEDLINE; 97175555.
RA   Mitchison M., Bulach D.M., Vinh T., Rajakumar K., Faine S., Adler B.;
RT   "Identification and characterization of the dTDP-rhamnose biosynthesis
RT   and transfer genes of the lipopolysaccharide-related rfb locus in
RT   Leptospira interrogans serovar Copenhageni.";
RL   J. Bacteriol. 179:1262-1267(1997).
DR   EMBL; U61226; AAB47845.1; -.
DR   INTERPRO; IPR001173; -.
DR   PFAM; PF00535; Glycos_transf_2; 1.
SQ   SEQUENCE   303 AA;  35272 MW;  1D3618DF517FB52D CRC64;

  Query Match             76.1%;  Score 35;  DB 2;  Length 303;
  Best Local Similarity   77.8%;  Pred. No. 28;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLNL 9
           | |||:|||
Db      22 YKNVQNLNL 30

RESULT   5
Q9S4F5
ID   Q9S4F5      PRELIMINARY;      PRT;   303 AA.
AC   Q9S4F5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HYPOTHETICAL 35.3 KDA PROTEIN.
OS   Leptospira interrogans.
OC   Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   De la Pena-Moctezuma A., Bulach D.M., Kalambaheti T., Adler B.;
RT   "Comparative analysis of the LPS biosynthetic loci of the genetic
RT   subtypes of serovar Hardjo: Leptospira interrogans Subtype
RT   Hardjoprajitno and Leptospira borgpetersenii Subtype Hardjobovis.";
RL   FEMS Microbiol. Lett. 177:319-326(1999).
DR   EMBL; AF144879; AAD52191.1; -.
```

```
DR   INTERPRO; IPR001173; -.
DR   PFAM; PF00535; Glycos_transf_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   303 AA;  35284 MW;  1D27E8DF517FB52D CRC64;


  Query Match             76.1%;  Score 35;  DB 2;  Length 303;
  Best Local Similarity   77.8%;  Pred. No. 28;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLNVQDLNL 9
                | |||:|||
Db         22 YKNVQNLNL 30


RESULT   6
Q9V496
ID   Q9V496      PRELIMINARY;      PRT;  2307 AA.
AC   Q9V496;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   RFABP PROTEIN.
GN   RFABP.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE; 20196006.
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003846; AAF59387.1; -.
DR   FLYBASE; FBgn0016724; RfaBp.
DR   INTERPRO; IPR001747; -.
DR   PFAM; PF01347; Vitellogenin_N; 2.
SQ   SEQUENCE   2307 AA;  255889 MW;  61A5F749A300B2E3 CRC64;


  Query Match             76.1%;  Score 35;  DB 5;  Length 2307;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 NVQDLNL 9
                |||||||
Db        726 NVQDLNL 732


RESULT   7
Q94907
ID   Q94907      PRELIMINARY;      PRT;  3351 AA.
AC   Q94907;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RETINOID- AND FATTY ACID-BINDING GLYCOPROTEIN.
GN   RFABP.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON;
RX   MEDLINE; 96355402.
RA   Kutty R.K., Kutty G., Kambadur R., Duncan T., Koonin E.V.,
RA   Rodriguez I.R., Odenwald W.F., Wiggert B.;
RT   "Molecular characterization and developmental expression of a
RT   retinoid-and fatty acid-binding glycoprotein from Drosophila. A
RT   putative lipophorin.";
RL   J. Biol. Chem. 271:20641-20649(1996).
DR   EMBL; U62892; AAC47284.1; -.
DR   FLYBASE; FBgn0016724; RfaBp.
DR   INTERPRO; IPR001747; -.
DR   INTERPRO; IPR001846; -.
DR   PFAM; PF00094; vwd; 1.
DR   PFAM; PF01347; Vitellogenin_N; 1.
SQ   SEQUENCE   3351 AA;  372640 MW;  D4ABFA99D02A17C8 CRC64;


  Query Match             76.1%;  Score 35;  DB 5;  Length 3351;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 NVQDLNL 9
                |||||||
Db        726 NVQDLNL 732


RESULT   8
O16355
ID   O16355      PRELIMINARY;      PRT;   421 AA.
AC   O16355;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   F44C8.9 PROTEIN.
GN   F44C8.9.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
```

```
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Dante M., Kramer J.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF016438; AAB65896.1; -.
DR   HSSP; P20393; 1A6Y.
DR   INTERPRO; IPR001628; -.
DR   PFAM; PF00105; zf-C4; 2.
DR   PRINTS; PR00047; STROIDFINGER.
SQ   SEQUENCE   421 AA;  48563 MW;  EE29AFA49DFF6A85 CRC64;

  Query Match             73.9%;  Score 34;  DB 5;  Length 421;
  Best Local Similarity   75.0%;  Pred. No. 63;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LNVQDLNL 9
            |||||:|:
Db      277 LNVQDINV 284


RESULT   9
Q44125
ID   Q44125      PRELIMINARY;      PRT;   898 AA.
AC   Q44125;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   TRANSFERRIN BINDING PROTEIN 1.
GN   TBPA.
OS   Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Actinobacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H49;
RX   MEDLINE; 96036198.
RA   Gonzalez G.C., Yu R.H., Rosteck P.R. Jr., Schryvers A.B.;
RT   "Sequence, genetic analysis, and expression of Actinobacillus
RT   pleuropneumoniae transferrin receptor genes.";
RL   Microbiology 141:2405-2416(1995).
DR   EMBL; U16017; AAC43485.1; -.
DR   INTERPRO; IPR000531; -.
DR   PFAM; PF00593; TonB_boxC; 1.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
SQ   SEQUENCE   898 AA;  102187 MW;  4E49988810A3752A CRC64;

  Query Match             73.9%;  Score 34;  DB 2;  Length 898;
  Best Local Similarity   66.7%;  Pred. No. 1.4e+02;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLNL 9
            | ||||: |
Db      709 YFNVQDIKL 717
```

```
RESULT  10
Q44158
ID   Q44158      PRELIMINARY;      PRT;   898 AA.
AC   Q44158;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   TRANSFERRIN RECEPTOR.
GN   TFBB.
OS   Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
OG   Plasmid pTF205/023.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Actinobacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AP205;
RA   Wilke M.;
RL   Thesis (1995), Institut fuer Mikrobiologie,
RL   Tieraerztliche Hochschule Hannover, Germany.
DR   EMBL; Z54191; CAA90896.1; -.
DR   INTERPRO; IPR000531; -.
DR   PFAM; PF00593; TonB_boxC; 1.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW   Plasmid.
SQ   SEQUENCE   898 AA;  102217 MW;  AD627D5BC0B7E6E2 CRC64;

  Query Match             73.9%;  Score 34;  DB 2;  Length 898;
  Best Local Similarity   66.7%;  Pred. No. 1.4e+02;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLNL 9
            | ||||: |
Db      709 YFNVQDIKL 717


RESULT  11
Q44127
ID   Q44127      PRELIMINARY;      PRT;   931 AA.
AC   Q44127;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   TRANSFERRIN BINDING PROTEIN 1.
GN   TBPA.
OS   Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Actinobacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H171;
RX   MEDLINE; 96036198.
RA   Gonzalez G.C., Yu R.H., Rosteck P.R. Jr., Schryvers A.B.;
RT   "Sequence, genetic analysis, and expression of Actinobacillus
RT   pleuropneumoniae transferrin receptor genes.";
RL   Microbiology 141:2405-2416(1995).
DR   EMBL; U16019; AAC43487.1; -.
DR   INTERPRO; IPR000531; -.
DR   PFAM; PF00593; TonB_boxC; 1.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
SQ   SEQUENCE   931 AA;  106369 MW;  CE88D077D03C54F9 CRC64;

  Query Match             73.9%;  Score 34;  DB 2;  Length 931;
  Best Local Similarity   66.7%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLNL 9
            | ||||: |
Db      742 YFNVQDIKL 750


RESULT  12
```

```
Q44166
ID   Q44166      PRELIMINARY;      PRT;   931 AA.
AC   Q44166;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   TBP1 GENE PRECURSOR.
GN   TBP1.
OS   Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Actinobacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 27088;
RX   MEDLINE; 96207589.
RA   Daban M., Medrano A., Querol E.;
RT   "Cloning, sequencing and expression of the transferrin-binding protein
RT   1 gene from Actinobacillus pleuropneumoniae.";
RL   Biochem. J. 315:257-264(1996).
DR   EMBL; Z49708; CAA89810.1; -.
DR   INTERPRO; IPR000531; -.
DR   PFAM; PF00593; TonB_boxC; 1.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW   Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    931       TBP1 GENE.
SQ   SEQUENCE   931 AA;  106326 MW;  D777366E2FDA54F9 CRC64;

  Query Match             73.9%;  Score 34;  DB 2;  Length 931;
  Best Local Similarity   66.7%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLNL 9
            | ||||: |
Db      742 YFNVQDIKL 750


RESULT  13
O77390
ID   O77390      PRELIMINARY;      PRT;   442 AA.
AC   O77390;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   PFC0720W PROTEIN.
GN   PFC0720W.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RA   Murphy L., Lawson D., Barrell B.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z98551; CAB11146.1; -.
SQ   SEQUENCE   442 AA;  52477 MW;  465780FA491DCC52 CRC64;

  Query Match             71.7%;  Score 33;  DB 5;  Length 442;
  Best Local Similarity   66.7%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLNL 9
            | || |||:
Db      374 YQNVDDLNI 382


RESULT  14
Q9XMT9
ID   Q9XMT9      PRELIMINARY;      PRT;   442 AA.
AC   Q9XMT9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   NADH DEHYDROGENASE SUBUNIT 7 (EC 1.6.5.3).
GN   NAD7.
OS   Tetrahymena pyriformis.
OG   Mitochondrion.
OC   Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC   Tetrahymenina; Tetrahymena.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Edqvist J., Burger G., Gray M.W.;
RT   "Expression of mitochondrial protein-coding genes in Tetrahymena
RT   pyriformis.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Burger G., Zhu Y., Littlejohn T.G., Greenwood S.J., Schnare M.N.,
RA   Lang B.F., Gray M.W.;
RT   "Complete sequence, gene content and organization of the mitochondrial
RT   genome of Tetrahymena pyriformis. Comparison with Paramecium aurelia
RT   mitochondrial DNA.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF160864; AAD41924.1; -.
DR   INTERPRO; IPR001135; -.
DR   PFAM; PF00346; complex1_49Kd; 2.
DR   PROSITE; PS00535; COMPLEX1_49K; 1.
KW   Oxidoreductase; Mitochondrion.
SQ   SEQUENCE   442 AA;  51109 MW;  8D8CF2AB2E4D24DF CRC64;

  Query Match             71.7%;  Score 33;  DB 8;  Length 442;
  Best Local Similarity   66.7%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLNL 9
            |||  |||:
Db      336 YLNKNDLNI 344


RESULT  15
Q9SJK3
ID   Q9SJK3      PRELIMINARY;      PRT;   997 AA.
AC   Q9SJK3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ARGONAUTE (AGO1)-LIKE PROTEIN.
GN   AT2G27880.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC   Brassicaceae; Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL   Nature 402:761-768(1999).
DR   EMBL; AC006929; AAD21514.1; -.
SQ   SEQUENCE   997 AA;  111134 MW;  DA93A1446C422F31 CRC64;

  Query Match             71.7%;  Score 33;  DB 10;  Length 997;
  Best Local Similarity   62.5%;  Pred. No. 2.5e+02;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLN 8
            :||::|||
```

```
Db     368 FLNIRDLN 375


Search completed: December 16, 2000, 02:01:01
Job time: 5935 sec
```

```
                        GenCore version 4.5
               Copyright (c) 1993 - 2000  Compugen Ltd.


OM protein - protein search, using sw model

Run on:         December 16, 2000, 00:51:15 ; Search time 108.84 Seconds
                                              (without alignments)
                                              2.827 Million cell updates/sec

Title:          US-09-529-121-9
Perfect score:  46
Sequence:       1 YLNVQDLNL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters:      268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_36:*
                 1:  /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
                 2:  /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
                 3:  /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
                 4:  /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
                 5:  /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
                 6:  /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
                 7:  /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
                 8:  /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
                 9:  /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
                 10:  /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
                 11:  /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
                 12:  /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
                 13:  /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
                 14:  /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
                 15:  /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
                 16:  /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
                 17:  /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
                 18:  /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
                 19:  /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
                 20:  /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
                 21:  /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                    %
Result              Query
  No.    Score      Match Length DB  ID                      Description
 --------------------------------------------------------------------------------
    1       46      100.0    240  20  Y28784                 Thymidylate kinase
    2       34       73.9    931  17  W04867                 Transferrin bindin
    3       33       71.7     43  20  Y73970                 Human prostate tum
    4       33       71.7   1464  17  R88469                 Feline infectious
    5       32       69.6    545  21  Y88391                 Tetrahydrocannabin
    6       32       69.6   2496  18  W36093                 Mutant YLR087c pro
    7       32       69.6   2942  18  W22050                 Saccharomyces cere
    8       31       67.4    528  20  W81457                 Canine ICAM-R poly
    9       31       67.4    528  21  Y50760                 Canine ICAM-R prot
   10       31       67.4    877  19  W59040                 Barley alpha-gluco
   11       31       67.4    976  18  W13408                 Arabidopsis thalia
   12       31       67.4    991  16  R80096                 Black widow spider
```

```
   13       31       67.4   1214  16  R80097                 Black widow spider
   14       31       67.4   1220  21  Y78972                 RPP1-WsB amino aci
   15       31       67.4   1430  20  Y31949                 Plasmodium falcipa
   16       30       65.2    185  15  R56968                 Human phosphatase
   17       30       65.2    185  18  W35330                 Human cdc25B vacci
   18       30       65.2    221  20  W87735                 RSPaV strain RSP15
   19       30       65.2    282  19  W38554                 S. pneumoniae form
   20       30       65.2    370  21  Y99428                 Human PRO1431 (UNQ
   21       30       65.2    511  19  W98255                 H. pylori GHPO 155
   22       30       65.2    511  19  W71519                 Helicobacter polyp
   23       30       65.2    623  15  R49578                 Sequence of bovine
   24       30       65.2    626  18  W16398                 Human neuron-deriv
   25       30       65.2    628  17  R92057                 Apoptopic cerebral
   26       30       65.2    780  20  Y35502                 Amino acid sequenc
   27       29       63.0     25  20  Y06845                 H. felis ORF1 prot
   28       29       63.0     33  19  W35039                 Water soluble beta
   29       29       63.0     60  18  W30353                 Fragment of glial-
   30       29       63.0     94  20  Y16719                 WO9914235 Seq ID N
   31       29       63.0     94  20  Y16666                 WO9914235 Seq ID N
   32       29       63.0    103  18  W15745                 (Gly33-Ile134) tru
   33       29       63.0    104  18  W15744                 (Arg32-Ile134) tru
   34       29       63.0    114  18  W15743                 (Pro23-Lys37delAsn
   35       29       63.0    130  16  R79096                 Human Fas ligand C
   36       29       63.0    133  16  R79376                 Human ATF-2.  Homo
   37       29       63.0    133  18  W30069                 Human glial-derive
   38       29       63.0    134  18  W32106                 Human partial glia
   39       29       63.0    134  18  W18052                 Mature human glial
   40       29       63.0    134  18  W18058                 Mature human glial
   41       29       63.0    134  18  W22027                 Glial cell derived
   42       29       63.0    134  18  W23782                 Human mature glial
   43       29       63.0    134  18  W15706                 Glial cell line-de
   44       29       63.0    134  20  Y16658                 WO9914235 Seq ID N
   45       29       63.0    134  20  Y16659                 WO9914235 Seq ID N


                                ALIGNMENTS


RESULT   1
Y28784
ID   Y28784 standard; protein; 240 AA.
XX
AC   Y28784;
XX
DT   02-NOV-1999  (first entry)
XX
DE   Thymidylate kinase-1.
XX
KW   Thymidylate kinase; TmpK; AIDS; prodrug; AZT; 3'-azido-3-deoxythymidine;
KW   viral replication; DNA chain termination; AZT activation pathway;
KW   AZT-MP; AZT-monophosphate; AZT-DP; AZT-diphosphate; anticancer;
KW   antiviral activity; therapeutic analog.
XX
OS   Unidentified.
XX
PN   WO9941404-A2.
XX
PD   19-AUG-1999.
XX
PF   12-FEB-1999;   99WO-EP00945.
XX
PR   13-FEB-1998;   98EP-0102546.
XX
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Goody RS,  Konrad M,  Lavie A,  Reinstein J,  Schlichting I;
XX
DR   WPI; 1999-508654/42.
XX
PT   Producing kinases with increased activity on nucleoside and
PT   nucleotide analogs, used to improve conversion of prodrugs, e.g.
PT   AZT, to active form
XX
```

```
PS   Claim 7; Page 68-69; 84pp; English.
XX
CC   The present sequence is a thymidylate kinase. TmpK is involved in the
CC   activation of the AIDS prodrug; AZT. AZT-TP inhibits viral replication
CC   by DNA chain termination. TmpK which catalyses the second
CC   phosphorylation step, from the monophosphate (AZT-MP) to the diphosphate
CC   (AZT-DP), is the rate limiting enzyme in the AZT activation pathway.
CC   Increasing the kinase activity of TmpK results in higher concentrations
CC   of the active form of the therapeutic analog especially AZT-triphosphate
CC   (AZT = 3'-azido-3-deoxythymidine) and thus a greater therapeutic effect,
CC   while reducing the concentration of the toxic monophosphate
CC   intermediate. This has anticancer and antiviral activities.
XX
SQ   Sequence   240 AA;


  Query Match             100.0%;  Score 46;  DB 20;  Length 240;
  Best Local Similarity   100.0%;  Pred. No. 0.1;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLNL 9
            |||||||||
Db      206 ylnvqdlnl 214


RESULT   2
W04867
ID   W04867 standard; Protein; 931 AA.
XX
AC   W04867;
XX
DT   19-JAN-1997  (first entry)
XX
DE   Transferrin binding protein 1 gene.
XX
KW   Transferrin binding protein 1; Tbp1; vaccine; antibody; diagnosis;
KW   swine pleuropneumonia.
XX
OS   Actinobacillus pleuropneumoniae strain 1371 serotype 1.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /label= Sig_peptide
XX
PN   EP733708-A2.
XX
PD   25-SEP-1996.
XX
PF   21-MAR-1996;   96EP-0870033.
XX
PR   24-MAR-1995;   95ES-0000592.
XX
PA   (HIPR-) LAB HIPRA SA.
XX
PI   Daban M,  Espuna E,  Medrano A,  Querol E;
XX
DR   WPI; 1996-427056/43.
DR   N-PSDB; T38071.
XX
PT   Actinobacillus pleuropneumoniae transferrin binding protein 1 - for
PT   production of antibodies useful diagnostically and in universal
PT   vaccine against porcine pleuropneumonia
XX
PS   Claim 12; Page 10-17; 22pp; English.
XX
CC   Transferrin binding protein 1 (Tbp1) (W04867) of Actinobacillus
CC   pleuropneumoniae (App) strain 1370, derived from strain Hpn-1
CC   (ATCC 27088) of serotype 1, is probably a transmembrane protein
CC   that serves as a channel for transport of iron across the outer
CC   membrane.  Its amino acid sequence was deduced from a gene
CC   (T38071) obtd. from App genomic DNA.  Recombinant Tbp1, or its
CC   antigenic fragments, can be produced in transformed host cells.
CC   It is used to formulate vaccines against porcine pleuropneumonia,
```

```
CC   to prepare antibodies (useful for serotherapy) and to prepare
CC   diagnostic reagents.
XX
SQ   Sequence   931 AA;


  Query Match             73.9%;  Score 34;  DB 17;  Length 931;
  Best Local Similarity   66.7%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLNL 9
            | ||||: |
Db      742 yfnvqdikl 750


RESULT   3
Y73970
ID   Y73970 standard; Protein; 43 AA.
XX
AC   Y73970;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Human prostate tumor EST fragment derived protein #157.
XX
KW   Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW   treatment.
XX
OS   Homo sapiens.
XX
PN   DE19820190-A1.
XX
PD   04-NOV-1999.
XX
PF   28-APR-1998;   98DE-1020190.
XX
PR   28-APR-1998;   98DE-1020190.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-621386/54.
DR   N-PSDB; Z52909.
XX
PT   New human nucleic acid sequences from pancreatic tumors, and related
PT   proteins  -
XX
PS   Claim 23; Page 375; 502pp; German.
XX
CC   This invention describes novel polypeptides and their encoding nucleic
CC   acids derived from human pancreatic tumor tissue which have cytostatic
CC   activity. The sequences are also useful in producing pharmaceutical
CC   compositions for treatment of pancreatic tumors. Y73814-Y74252 represent
CC   protein fragments encoded by the human pancreatic tumor cDNA library
CC   derived expressed sequence tag (EST) sequences represented in
CC   Z52858-Z53014.
XX
SQ   Sequence   43 AA;


  Query Match             71.7%;  Score 33;  DB 20;  Length 43;
  Best Local Similarity   66.7%;  Pred. No. 5.4;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLNL 9
            ||| ||::|
Db        7 ylnfqdvsl 15


RESULT   4
R88469
ID   R88469 standard; Protein; 1464 AA.
```

```
XX
AC   R88469;
XX
DT   14-AUG-1996  (first entry)
XX
DE   Feline infectious peritonitis 1 virus spike protein.
XX
KW   Feline infectious peritonitis 1 virus; FIPV-I; spike protein;
KW   vaccine; prevention; treatment.
XX
OS   Feline infectious peritonitis 1 virus.
XX
PN   JP07327683-A.
XX
PD   19-DEC-1995.
XX
PF   10-JUN-1994;   94JP-0129300.
XX
PR   10-JUN-1994;   94JP-0129300.
XX
PA   (KITA ) KITASATO KENKYUSHO SH.
XX
DR   WPI; 1996-072341/08.
DR   N-PSDB; T10166.
XX
PT   DNA encoding feline infectious peritonitis I virus spike protein -
PT   used in a vaccine for prevention and treatment of FIPV-I infection
XX
PS   Claim 1; Page 14-17; 23pp; Japanese.
XX
CC   This sequence represents the feline infectious peritonitis 1 virus
CC   (FIPV-I) spike protein.  The FIPV-I spike protein may be used in the
CC   production of a vaccine for the prevention and treatment of FIPV-I
CC   infection.  The spike protein may be produced by transforming a host
CC   cell with the spike protein DNA and expressing the sequence such
CC   that the spike protein can be isolated.
XX
SQ   Sequence   1464 AA;


  Query Match             71.7%;  Score 33;  DB 17;  Length 1464;
  Best Local Similarity   55.6%;  Pred. No. 2.8e+02;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLNL 9
           |:| |:||:
Db      86 yintqnlni 94


RESULT   5
Y88391
ID   Y88391 standard; Protein; 545 AA.
XX
AC   Y88391;
XX
DT   27-JUL-2000  (first entry)
XX
DE   Tetrahydrocannabinolic acid synthase amino acid sequence.
XX
KW   Tetrahydrocannabinolic acid synthase; THCA; anaesthesia; pain-killer;
KW   intraocular pressure lowering; anti-inflammatory treatment.
XX
OS   Cannabis sativa.
XX
PN   JP2000078979-A.
XX
PD   21-MAR-2000.
XX
PF   04-SEP-1998;   98JP-0251667.
XX
PR   04-SEP-1998;   98JP-0251667.
XX
PA   (TAIS ) TAISHO PHARM CO LTD.
XX
DR   WPI; 2000-285930/25.
DR   N-PSDB; A13340, A13341.
XX
PT   New tetrahydrocannabinolic acid synthase gene  -
XX
PS   Claim 2; Page 8-10; 17pp; Japanese.
XX
CC   This sequence represents the Cannabis sativa tetrahydrocannabinolic acid
CC   (THCA) synthase amino acid sequence. The THCA gene sequence can be used
CC   to produce an expression vector containing the gene. Cells can be
CC   transformed using the vector, so that they produce a protein with THCA
CC   synthase activity. THCA can be used for anaesthesia, pain-killing,
CC   intraocular pressure lowering and anti-inflammatory treatment.
XX
SQ   Sequence   545 AA;

  Query Match             69.6%;  Score 32;  DB 21;  Length 545;
  Best Local Similarity   66.7%;  Pred. No. 1.4e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLNL 9
           ||| :||:|
Db     481 ylnyrdldl 489


RESULT   6
W36093
ID   W36093 standard; Protein; 2496 AA.
XX
AC   W36093;
XX
DT   25-MAR-1998  (first entry)
XX
DE   Mutant YLR087c protein from cold sensitive yeast strain.
XX
KW   Mutant; cold sensitive; yeast; complementation; breadmaking; dough;
KW   sugar; leavening agent; brewing; winemaking; truncation.
XX
OS   Saccharomyces cerevisiae.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 527
FT                   /note= "encoded by GCR"
FT   Misc-difference 618
FT                   /note= "encoded by GAY"
FT   Misc-difference 640
FT                   /note= "encoded by TCY"
FT   Misc-difference 1103
FT                   /note= "encoded by AGR"
FT   Misc-difference 1582
FT                   /note= "encoded by GAR"
XX
PN   WO9728693-A1.
XX
PD   14-AUG-1997.
XX
PF   07-FEB-1997;   97WO-FR00254.
XX
PR   08-FEB-1996;   96FR-0001562.
XX
PA   (LESA ) LESAFFRE & CIE.
XX
PI   Colavizza D,  Loiez A,  Wadoux I;
XX
DR   WPI; 1997-414988/38.
DR   N-PSDB; T94548.
XX
PT   New strains of bread-making yeast with low fermentative activity at
PT   low temperature - allows production of doughs that can be stored
PT   cold for many hours before final baking
XX
```

```
PS   Example 4; Page 59-64; 76pp; French.
XX
CC   This is the amino acid sequence of a mutant protein (designated YLR087c)
CC   isolated from a cold sensitive strain of Saccharomyces cerevisiae.  The
CC   gene sequence was isolated from the clone YCp50-10.39 which was able to
CC   complement the cold sensitive strain HL13.2.30.  This protein is a
CC   truncated mutation as compared to the wild type sequence.  The wild type
CC   gene contains a open reading frame of 8874 bases encoding a protein of
CC   2958 amino acids.  The mutant gene has an open reading frame of 7488
CC   bases encoding a protein of 2496 amino acids.  Yeast strains containing
CC   the YLR087c gene are used, fresh or dried, in modified breadmaking
CC   processes, i.e. in processes where the delay between mixing the dough
CC   and baking exceeds 6 hr.  Particularly they are used to make French-style
CC   bread (no added sugar) or breads with sugar content below 5%. The strains
CC   can also be used to make mixed yeast/bacteria acidic leavening agents.
CC   The strains can also be used in brewing and winemaking as a
CC   cold-sensitive phenotype.
XX
SQ   Sequence   2496 AA;


  Query Match             69.6%;  Score 32;  DB 18;  Length 2496;
  Best Local Similarity   66.7%;  Pred. No. 7.9e+02;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLNVQDLNL 9
              |||| | |:
Db        642 ylnvndhni 650


RESULT   7
W22050
ID   W22050 standard; Protein; 2942 AA.
XX
AC   W22050;
XX
DT   26-FEB-1998  (first entry)
XX
DE   Saccharomyces cerevisiae protein encoded by novel CSF-1 gene.
XX
KW   Saccharomyces cerevisiae; cold sensitivity of fermentability; CSF-1;
KW   low temperature sensitivity; yeast; bread; ethanol; fermentation.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO9724442-A1.
XX
PD   10-JUL-1997.
XX
PF   27-DEC-1996;   96WO-JP03862.
XX
PR   28-DEC-1995;   95JP-0343700.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Kawasaki H,  Kikuchi Y,  Ouchi K,  Tokai M;
XX
DR   WPI; 1997-363678/33.
DR   N-PSDB; T74989.
XX
PT   Protein for complementing low temperature sensitivity in
PT   fermentation - can be inactivated in yeast and used for producing
PT   bread and ethanol
XX
PS   Claim 3; Pages 21-40; 48pp; Japanese.
XX
CC   This is a protein capable of complementing a mutation showing low
CC   temperature sensitivity in fermentation. This is encoded by a novel
CC   genomic DNA CSF-1 (cold sensitivity of fermentability). The Saccharomyces
CC   cerevisiae YHK 1243 has an inactivated gene encoding this new protein or
CC   has the novel DNA sequence contained in its chromosome. This protein can
CC   be used in dough to produce bread, and to produce ethanol.
XX
SQ   Sequence   2942 AA;

  Query Match             69.6%;  Score 32;  DB 18;  Length 2942;
  Best Local Similarity   66.7%;  Pred. No. 9.5e+02;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLNVQDLNL 9
              |||| | |:
Db        642 ylnvndhni 650


RESULT   8
W81457
ID   W81457 standard; Protein; 528 AA.
XX
AC   W81457;
XX
DT   17-FEB-1999  (first entry)
XX
DE   Canine ICAM-R polypeptide sequence.
XX
KW   Intercellular adhesion molecule polypeptide; ICAM-R; humanised; ICR 1.1;
KW   ICR 8.1; monoclonal antibody; therapeutic; inflammatory; asthma; tumour;
KW   graft-versus-host disease; viral infection; toxin; radionuclide;
KW   neovascularisation site; canine.
XX
OS   Canis sp.
XX
PN   US5837822-A.
XX
PD   17-NOV-1998.
XX
PF   07-JUN-1995;   95US-0487113.
XX
PR   07-JUN-1995;   95US-0487113.
PR   27-JAN-1992;   92US-0827689.
PR   26-MAY-1992;   92US-0889724.
PR   05-JUN-1992;   92US-0894061.
PR   22-JAN-1993;   93US-0009266.
PR   26-JAN-1993;   93WO-US00787.
PR   05-AUG-1993;   93US-0102852.
XX
PA   (ICOS-) ICOS CORP.
XX
PI   Gallatin WM,  Vazeux R;
XX
DR   WPI; 1999-023535/02.
DR   N-PSDB; V69206.
XX
PT   Humanised antibodies specific for intercellular adhesion molecule
PT   polypeptide - useful for therapeutic or diagnostic purposes
XX
PS   Example 31; Columns 157-160; 116pp; English.
XX
CC   The invention relates to humanised ICR 1.1 and ICR 8.1 antibodies
CC   targeted to the human intercellular adhesion molecule polypeptide
CC   (ICAM-R) polypeptide. Antibodies specific for ICAM's are potentially
CC   useful as therapeutic compounds, for treating e.g. immune-mediated
CC   inflammatory conditions (e.g. graft-versus-host disease), asthma,
CC   tumours or viral infections. Monoclonal antibodies specific for ICAM-R,
CC   or their conjugates formed with e.g. toxins or radionuclides are useful
CC   for therapeutically targeting or detecting neovascularisation sites.
CC   This represents the amino acid sequence of the longest canine ICAM-R
CC   clone, isolated using human ICAM-R sequences.
XX
SQ   Sequence   528 AA;


  Query Match             67.4%;  Score 31;  DB 20;  Length 528;
  Best Local Similarity   75.0%;  Pred. No. 2.2e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy       2 LNVQDLNL 9
           :|||| ||
Db     461 mnvqdrnl 468


RESULT   9
Y50760
ID   Y50760 standard; Protein; 528 AA.
XX
AC   Y50760;
XX
DT   16-FEB-2000  (first entry)
XX
DE   Canine ICAM-R protein #1.
XX
KW   ICAM-R; canine; intercellular adhesion molecule; phosphorylation;
KW   protein kinase C; modulator.
XX
OS   Canis familiaris.
XX
PN   US5989843-A.
XX
PD   23-NOV-1999.
XX
PF   27-SEP-1996;   96US-0720420.
XX
PR   27-JAN-1992;   92US-0827689.
PR   26-MAY-1992;   92US-0889724.
PR   05-JUN-1992;   92US-0894061.
PR   22-JAN-1993;   93US-0009266.
PR   26-JAN-1993;   93WO-US00787.
PR   05-AUG-1993;   93US-0102852.
PR   07-JUN-1995;   95US-0487113.
XX
PA   (ICOS-) ICOS CORP.
XX
PI   Gallatin WM,  Vazeux R;
XX
DR   WPI; 2000-022778/02.
DR   N-PSDB; Z24362.
XX
PT   Identifying modulators of protein kinase C phosphorylation of human
PT   intercellular adhesion molecule polypeptide -
XX
PS   Example 34; Column 163-166; 122pp; English.
XX
CC   This invention describes a novel method for identifying a compound that
CC   modulates phosphorylation of human intercellular adhesion molecule
CC   polypeptide (ICAM-R) by protein kinase C isoform. The method comprises:
CC   (a) exposing a purified peptide consisting of the cytoplasmic domain of
CC   ICAM-R to protein kinase C isoform and labeled adenosine triphosphate in
CC   the presence and absence of a test compound; (b) measuring labeled
CC   phosphate transferred to the peptide; and (c) identifying a test compound
CC   that affects transfer of the labeled phosphate as a modulator compound.
CC   The method is useful for identifying compounds that modulate the
CC   phosphorylation of human intercellular adhesion molecule polypeptide
CC   which might form the basis for the development of therapeutic and
CC   diagnostic agents. This sequence represents a canine ICAM-R protein
CC   which is described in the method of the invention.
XX
SQ   Sequence   528 AA;


  Query Match             67.4%;  Score 31;  DB 21;  Length 528;
  Best Local Similarity   75.0%;  Pred. No. 2.2e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LNVQDLNL 9
           :|||| ||
Db     461 mnvqdrnl 468


RESULT  10
W59040
ID   W59040 standard; Protein; 877 AA.
XX
AC   W59040;
XX
DT   07-AUG-1998  (first entry)
XX
DE   Barley alpha-glucosidase protein.
XX
KW   Alpha-glucosidase; barley; starch grain hydrolysis; alpha-amylase;
KW   glucoamylase; industry; germplasm; hydrolytic enzyme.
XX
OS   Hordeum vulgare.
XX
PN   US5763252-A.
XX
PD   09-JUN-1998.
XX
PF   28-APR-1995;   95US-0430925.
XX
PR   28-APR-1995;   95US-0430925.
XX
PA   (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI   Skadsen RW,  Tibbot BK;
XX
DR   WPI; 1998-347329/30.
DR   N-PSDB; V11736.
XX
PT   DNA encoding barley alpha-glucosidase protein - useful for producing
PT   recombinant protein to increase rate of starch grain hydrolysis when
PT   used with alpha amylase
XX
PS   Claim 1; Col 21-26; 19pp; English.
XX
CC   This sequence represents a novel barley alpha-glucosidase protein.
CC   Recombinant alpha-glucosidase can be used to increase the rate of starch
CC   grain hydrolysis when used together with alpha-amylase or can supplement
CC   glucoamylase in industrial starch hydrolysis systems. Useful DNA sequence
CC   characteristics from this enzyme can be identified which can be used as
CC   hybridisation probes for identifying germplasm with high levels of
CC   efficient hydrolytic enzymes.
XX
SQ   Sequence   877 AA;


  Query Match             67.4%;  Score 31;  DB 19;  Length 877;
  Best Local Similarity   85.7%;  Pred. No. 3.9e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLNVQDL 7
            |||| ||
Db     296 ylnvsdl 302


RESULT  11
W13408
ID   W13408 standard; Protein; 976 AA.
XX
AC   W13408;
XX
DT   10-JUN-1997  (first entry)
XX
DE   Arabidopsis thaliana plant morphogenesis regulatory protein.
XX
KW   Plant; morphogenesis; regulation; short; stem; alteration;
KW   inflorescence; extraneous; gene; expression; transformation;
KW   increase; control; form; length.
XX
OS   Arabidopsis thaliana.
XX
PN   JP09056382-A.
XX
```

```
PD   04-MAR-1997.
XX
PF   24-AUG-1995;   95JP-0216187.
XX
PR   24-AUG-1995;   95JP-0216187.
XX
PA   (MITS-) MITSUI GYOSAI SHOKUBUTSU BIO KENKYUSHO.
PA   (CHIK-) ZH CHIKYU KANKYO SANGYO GIJITSU KENKYU.
XX
DR   WPI; 1997-206629/19.
DR   N-PSDB; T62124.
XX
PT   DNA encoding plant morphogenesis regulatory protein - useful to
PT   yield plants with short stems or altered inflorescence
XX
PS   Claim 1; Pages 6-10; 17pp; Japanese.
XX
CC   The present sequence is an Arabidopsis thaliana plant
CC   morphogenesis regulatory protein (MRP), which can be used to yield
CC   a plant with, e.g. short stems or altered inflorescence. The MRP
CC   acts on a plant at a specific site for a specific period, and can
CC   therefore be used to regulate extraneous gene expression in a
CC   plant. The MRP's cDNA or genomic DNA can be used to transform a
CC   plant to increase its MRP expression, and therefore control the
CC   form (particularly stem length) of the plant.
XX
SQ   Sequence   976 AA;


  Query Match             67.4%;  Score 31;  DB 18;  Length 976;
  Best Local Similarity   75.0%;  Pred. No. 4.4e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LNVQDLNL 9
           ||: ||||
Db      73 lnlsdlnl 80


RESULT  12
R80096
ID   R80096 standard; Protein; 991 AA.
XX
AC   R80096;
XX
DT   16-MAR-1996  (first entry)
XX
DE   Black widow spider delta-latroinsectotoxin.
XX
KW   Black widow spider; insecticide; pesticide; delta-latroinsectotoxin;
KW   spider venom; neurotoxin; toxin.
XX
OS   Latrodectus mactans tredecimguttatus.
XX
PN   GB2288807-A.
XX
PD   01-NOV-1995.
XX
PF   24-APR-1995;   95GB-0008298.
XX
PR   27-APR-1994;   94GB-0008466.
XX
PA   (BRTE-) BRITISH TECHNOLOGY GROUP LTD.
XX
PI   Bell DR,  Dulubova I,  Galkina TG,  Grishin E,  Khvotchev MV;
PI   Krasnoperov V,  Pluzhnikov KA,  Shamotienko OG,  Usherwood PNR;
PI   Volkova T,  Galkina T,  Khovotchev MV;
XX
DR   WPI; 1995-360758/47.
DR   N-PSDB; T04687.
XX
PT   Polypeptide(s) expressed by truncated genes, esp. spider
PT   delta-latroinsectotoxin  - also related non toxic precursor
PT   polypeptide. isolated from Black Widow spider, useful as insecticide
```

```
XX
PS   Claim 10; Page 35-38; 62pp; English.
XX
CC   This protein may be expressed recombinantly in E. coli BL21 (
CC   DE3) cells transformed with pT7-7 vectors comprising the truncated
CC   form of the sequence.  The encoded protein is an insect-specific
CC   neurotoxin, delta-lactroinsectotoxin, which is useful as an oral or
CC   topical insecticide.  Expressing the truncated gene allows large-
CC   scale production of active toxin in bacteria, eliminating the need
CC   to extract it from spider venom.  This neurotoxin is harmless to
CC   mammals and plants.
XX
SQ   Sequence   991 AA;


  Query Match             67.4%;  Score 31;  DB 16;  Length 991;
  Best Local Similarity   71.4%;  Pred. No. 4.4e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LNVQDLN 8
           ||:||:|
Db     636 lniqdvn 642


RESULT  13
R80097
ID   R80097 standard; Protein; 1214 AA.
XX
AC   R80097;
XX
DT   16-MAR-1996  (first entry)
XX
DE   Black widow spider delta-latroinsectotoxin precursor protein.
XX
KW   Black widow spider; insecticide; pesticide; delta-latroinsectotoxin;
KW   spider venom; neurotoxin; toxin.
XX
OS   Latrodectus mactans tredecimguttatus.
XX
PN   GB2288807-A.
XX
PD   01-NOV-1995.
XX
PF   24-APR-1995;   95GB-0008298.
XX
PR   27-APR-1994;   94GB-0008466.
XX
PA   (BRTE-) BRITISH TECHNOLOGY GROUP LTD.
XX
PI   Bell DR,  Dulubova I,  Galkina TG,  Grishin E,  Khvotchev MV;
PI   Krasnoperov V,  Pluzhnikov KA,  Shamotienko OG,  Usherwood PNR;
PI   Volkova T,  Galkina T,  Khovotchev MV;
XX
DR   WPI; 1995-360758/47.
DR   N-PSDB; T04688.
XX
PT   Polypeptide(s) expressed by truncated genes, esp. spider
PT   delta-latroinsectotoxin  - also related non toxic precursor
PT   polypeptide. isolated from Black Widow spider, useful as insecticide
XX
PS   Claim 61; Page 43-47; 62pp; English.
XX
CC   This protein is an insect-specific neurotoxin, delta-
CC   lactroinsectotoxin, which is useful as an oral or topical
CC   insecticide.  This neurotoxin is harmless to mammals and plants.
XX
SQ   Sequence   1214 AA;


  Query Match             67.4%;  Score 31;  DB 16;  Length 1214;
  Best Local Similarity   71.4%;  Pred. No. 5.6e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       2 LNVQDLN 8
             ||:||:|
Db     664 lniqdvn 670


RESULT  14
Y78972
ID   Y78972 standard; Protein; 1220 AA.
XX
AC   Y78972;
XX
DT   05-JUN-2000  (first entry)
XX
DE   RPP1-WsB amino acid sequence.
XX
KW   Recognition of Peronospora parasitica 1; RPP1-WsB; downy mildew;
KW   biotrophic oomycete; chromosome 3; nucleotide binding-LRR protein;
KW   fungicide; transgenic plant.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200008189-A2.
XX
PD   17-FEB-2000.
XX
PF   09-AUG-1999;   99WO-GB02609.
XX
PR   07-AUG-1998;   98GB-0017278.
XX
PA   (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI   Jones JDG,  Beynon JL,  Parker JE,  Botella-Mesa MA;
XX
DR   WPI; 2000-205727/18.
DR   N-PSDB; Z98619, Z98620.
XX
PT   Peronospora parasitica resistance genes from Arabidopsis thaliana
PT   useful in transgenic plants having increased resistance to the fungus
PT   -
XX
PS   Claim 2; Fig 3G-3H; 80pp; English.
XX
CC   This sequence represents the protein sequence encoded by the Arabidopsis
CC   thaliana (Wassilewskija cultivar) Peronospora parasitica resistance
CC   RPP1-WsB gene. The Arabidopsis RPP1 locus has at least three functional
CC   genes designated RPP1-WsA, RPP1-WsB, and RPP1-WsC, that differ in their
CC   ability to detect four different P. parasitica isolates. Peronospora
CC   parasitica (downy mildew) is a biotrophic oomycete, and Arabidopsis RPP1
CC   genes encode polypeptides capable of recognising and activating a defence
CC   response in a plant in response to challenge with a P. parasitica
CC   isolate, or a P. parasitica Avr (avirulence) gene. The three RPP-1 genes
CC   encode functional proteins of the Nucleotide Binding (NB)-LRR resistance
CC   protein class. Each gene recognises different pathogen avirulence genes
CC   or alleles. Together the RPP1 genes comprise a complex resistance locus
CC   containing a functionally variable family of resistance genes found on
CC   the bottom arm of chromosome 3. The RPP1 nucleotide sequences can be used
CC   for identifying, cloning or amplifying RPP1 sequences. Transgenic plants
CC   containing the RPP1 sequences have increased resistance to the fungus
CC   P. parasitica. The methods are useful for influencing or affecting the
CC   degree of resistance of a plant to such a fungus.
XX
SQ   Sequence   1220 AA;


  Query Match             67.4%;  Score 31;  DB 21;  Length 1220;
  Best Local Similarity   55.6%;  Pred. No. 5.6e+02;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLNL 9
             | ||::||:
Db     611 yknveelni 619
```

```
RESULT  15
Y31949
ID   Y31949 standard; Protein; 1430 AA.
XX
AC   Y31949;
XX
DT   21-DEC-1999  (first entry)
XX
DE   Plasmodium falciparum cytoadherence gene protein CLAG2.
XX
KW   CLAG2; cytoadherence linked asexual gene; CLAG; erythrocyte;
KW   red blood cell; malaria; infection; therapy; vaccine.
XX
OS   Plasmodium falciparum.
XX
PN   WO9949048-A1.
XX
PD   30-SEP-1999.
XX
PF   25-MAR-1999;   99WO-AU00213.
XX
PR   25-MAR-1998;   98AU-0002580.
XX
PA   (MENZ-) MENZIES SCHOOL HEALTH RES.
XX
PI   Kemp DJ,  Trenholme KR,  Gardiner DL,  Holt DC,  Cowman AF;
XX
DR   WPI; 1999-591099/50.
XX
PT   New proteins useful for treatment of Plasmodium infections in humans,
PT   especially malaria  -
XX
PS   Claim 18; Page 136-145; 150pp; English.
XX
CC   This sequence represents the Plasmodium falciparum cytoadherence
CC   linked asexual gene 2 (CLAG2) protein that facilitates
CC   cytoadherence of erythrocytes parasitised with Plasmodium spp. to
CC   other cells.  It is encoded by a gene located on chromosome 2 of
CC   P. falciparum.  It is a member of a family of CLAG proteins (see
CC   also Y31945-48).  The identification of clag genetic species, and
CC   the products encoded by them, enables a range of therapeutic agents
CC   to be rationally designed and/or identified that are useful for the
CC   prophylaxis and treatment of disease conditions caused or
CC   exacerbated by infection with Plasmodium spp., e.g. malaria,
CC   especially in humans.
XX
SQ   Sequence   1430 AA;


  Query Match             67.4%;  Score 31;  DB 20;  Length 1430;
  Best Local Similarity   83.3%;  Pred. No. 6.7e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLNVQD 6
             |||:||
Db     645 ylniqd 650


Search completed: December 16, 2000, 00:51:16
Job time: 18776 sec
```

```
                    GenCore version 4.5
           Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 04:22:13 ; Search time 111.26 Seconds
                                              (without alignments)
                                              7.553 Million cell updates/sec

Title:           US-09-529-121-9
Perfect score:   46
Sequence:        1 YLNVQDLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters:      436

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       SPTREMBL_14:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                     SUMMARIES
                    %
Result            Query
  No.    Score    Match Length DB  ID                     Description
-------------------------------------------------------------------------------
     1      21     45.7      9  5  P82003                 P82003 bombyx mori
     2      18     39.1      9  4  Q9UCN5                 Q9ucn5 homo sapien
     3      17     37.0      8  2  P77556                 P77556 escherichia
     4      15     32.6      8  8  Q9XNP8                 Q9xnp8 boophilus m
     5      14     30.4      8  2  P72279                 P72279 rhodococcus
     6      14     30.4      9  1  Q50832                 Q50832 methanococc
     7      13     28.3      7  8  O98866                 O98866 spinacia ol
     8      13     28.3      8  8  Q35792                 Q35792 saccharomyc
     9      13     28.3      8 10   Q40659                 Q40659 oryza sativ
    10      13     28.3      9  2  Q48686                 Q48686 lactococcus
    11      13     28.3      9  4  Q14715                 Q14715 homo sapien
    12      13     28.3      9  4  O95953                 O95953 homo sapien
    13      13     28.3      9  9  Q9XJN0                 Q9xjn0 bacteriopha
    14      12     26.1      8  2  Q9RQ57                 Q9rq57 buchnera ap
    15      12     26.1      8  2  Q9RQ49                 Q9rq49 buchnera ap
    16      12     26.1      8  3  P87225                 P87225 saccharomyc
    17      12     26.1      8  4  Q15889                 Q15889 homo sapien
    18      12     26.1      8  4  Q9UMH9                 Q9umh9 homo sapien
    19      12     26.1      8  4  Q9UL56                 Q9ul56 homo sapien
    20      12     26.1      9  2  Q57328                 Q57328 aeromonas s
    21      12     26.1      9  2  Q44001                 Q44001 aeromonas e
    22      12     26.1      9  2  Q44377                 Q44377 aeromonas t
    23      12     26.1      9  2  Q44468                 Q44468 aeromonas v
    24      12     26.1      9  2  Q43928                 Q43928 aeromonas c
    25      12     26.1      9  4  Q9UQW0                 Q9uqw0 homo sapien
    26      12     26.1      9  4  Q9UQA3                 Q9uqa3 homo sapien
    27      12     26.1      9  5  O96417                 O96417 drosophila
    28      12     26.1      9 12   Q88612                 Q88612 middelburg
    29      12     26.1      9 12   Q89491                 Q89491 murine minu
    30      11     23.9      7 12   Q9YQ10                 Q9yq10 porcine tra
    31      11     23.9      8  2  Q51594                 Q51594 escherichia
    32      11     23.9      8  2  Q9ZEZ9                 Q9zez9 buchnera ap
    33      11     23.9      8  4  Q9UDZ4                 Q9udz4 homo sapien
    34      11     23.9      8  8  Q34909                 Q34909 locusta mig
    35      11     23.9      8  8  Q9T2W0                 Q9t2w0 saccharomyc
    36      11     23.9      8  9  Q37854                 Q37854 bacteriopha
    37      11     23.9      8 11   Q9QVF4                 Q9qvf4 rattus sp.
    38      11     23.9      8 12   Q83332                 Q83332 murine hepa
    39      11     23.9      8 13   P82082                 P82082 limnodynast
    40      11     23.9      9  2  Q9R7H9                 Q9r7h9 haemophilus
    41      11     23.9      9  4  P78484                 P78484 homo sapien
    42      11     23.9      9  4  Q16386                 Q16386 homo sapien
    43      11     23.9      9  6  Q9TRW2                 Q9trw2 oryctolagus
    44      11     23.9      9 12   Q84333                 Q84333 simian viru
    45      10     21.7      4 11   Q08433                 Q08433 rattus norv
    46      10     21.7      7  2  Q54248                 Q54248 streptomyce
    47      10     21.7      7  2  O07354                 O07354 synechococc
    48      10     21.7      7  5  Q9VYN9                 Q9vyn9 drosophila
    49      10     21.7      7 11   Q63480                 Q63480 rattus norv
    50      10     21.7      8  2  Q9ZIE9                 Q9zie9 neisseria m
    51      10     21.7      8  3  Q05403                 Q05403 saccharomyc
    52      10     21.7      8  4  Q9Y4J4                 Q9y4j4 homo sapien
    53      10     21.7      8  4  Q9Y4J3                 Q9y4j3 homo sapien
    54      10     21.7      8  5  Q94695                 Q94695 physarum po
    55      10     21.7      8  5  Q9UB13                 Q9ub13 albinaria h
    56      10     21.7      8 10   Q9SB24                 Q9sb24 nicotiana t
    57      10     21.7      8 11   Q62527                 Q62527 mus spretus
    58      10     21.7      8 13   P82083                 P82083 limnodynast
    59      10     21.7      9  2  Q9R9C4                 Q9r9c4 borrelia bu
    60      10     21.7      9  4  Q16276                 Q16276 homo sapien
    61      10     21.7      9  5  Q9VV82                 Q9vv82 drosophila
    62      10     21.7      9  8  Q31653                 Q31653 anser caeru
    63      10     21.7      9  8  O78337                 O78337 caloglossa
    64      10     21.7      9  8  Q9TLD0                 Q9tld0 bostrychia
    65      10     21.7      9  8  Q9TJ87                 Q9tj87 caloglossa
    66      10     21.7      9  8  Q9TJ85                 Q9tj85 caloglossa
    67      10     21.7      9  8  Q9T389                 Q9t389 caloglossa
    68      10     21.7      9  8  Q9T388                 Q9t388 caloglossa
    69      10     21.7      9  8  Q9T387                 Q9t387 bostrychia
    70      10     21.7      9 10   Q42452                 Q42452 triticum ae
    71      10     21.7      9 10   O81962                 O81962 caloglossa
    72      10     21.7      9 10   O81964                 O81964 caloglossa
    73      10     21.7      9 10   O81966                 O81966 caloglossa
    74      10     21.7      9 10   O81968                 O81968 caloglossa
    75      10     21.7      9 10   O82778                 O82778 caloglossa

                                   ALIGNMENTS


RESULT   1
P82003
ID   P82003     PRELIMINARY;      PRT;     9 AA.
AC   P82003;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   PROTHORACICOSTATIC PEPTIDE (PTSP).
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
```

```
RN   [1]
RP   SEQUENCE.
RC   STRAIN=C145 X N140; TISSUE=BRAIN;
RA   Hua Y.-J., Tanaka Y., Nakamura K.;
RT   "Identification of a prothoracicostatic peptide (PTSP) from the larval
RT   brain of the silkworm, Bombyx mori.";
RL   J. Biol. Chem. 0:0-0(1999).
CC   -!- FUNCTION: INHIBITS ECDYSTEROIDOGENESIS BY PROTHORACIC GLAND IN THE
CC       SILKWORM.
CC   -!- DEVELOPMENTAL STAGE: EARLY FIFTH INSTAR.
KW   Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1090 MW;  3878C5B4472AB6C3 CRC64;


  Query Match             45.7%;  Score 21;  DB 5;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 QDLN 8
          ||||
Db      3 QDLN 6


RESULT   2
Q9UCN5
ID   Q9UCN5      PRELIMINARY;      PRT;     9 AA.
AC   Q9UCN5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CELL-SURFACE HEPARIN/HEPARANSULFATE-BINDING PROTEIN PEPTIDE 2
DE   (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92291065.
RA   Raboudi N., Julian J., Rohde L.H., Carson D.D.;
RT   "Identification of cell-surface heparin/heparan sulfate-binding
RT   proteins of a human uterine epithelial cell line (RL95).";
RL   J. Biol. Chem. 267:11930-11939(1992).
SQ   SEQUENCE   9 AA;  1008 MW;  CB56D0544732C732 CRC64;


  Query Match             39.1%;  Score 18;  DB 4;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 LNVQ 5
          ||:|
Db      5 LNIQ 8


RESULT   3
P77556
ID   P77556      PRELIMINARY;      PRT;     8 AA.
AC   P77556;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   TRAY (FRAGMENT).
GN   TRAY.
OS   Escherichia coli.
OG   Plasmid IncFII R1.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ECOR11;
RX   MEDLINE; 96400908.
RA   Boyd E.F., Hill C.W., Rich S.M., Hartl D.L.;
```

```
RT   "Mosaic structure of plasmids from natural populations of Escherichia
RT   coli.";
RL   Genetics 143:1091-1100(1996).
DR   EMBL; U50661; AAC44245.1; -.
DR   EMBL; U50650; AAC44234.1; -.
DR   EMBL; U50651; AAC44235.1; -.
DR   EMBL; U50652; AAC44236.1; -.
DR   EMBL; U50653; AAC44237.1; -.
DR   EMBL; U50654; AAC44238.1; -.
DR   EMBL; U50655; AAC44239.1; -.
DR   EMBL; U50656; AAC44240.1; -.
DR   EMBL; U50657; AAC44241.1; -.
DR   EMBL; U50658; AAC44242.1; -.
DR   EMBL; U50659; AAC44243.1; -.
DR   EMBL; U50660; AAC44244.1; -.
KW   Plasmid.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  834 MW;  D335A5B0544735A1 CRC64;


  Query Match             37.0%;  Score 17;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 LNVQDL 7
          ||:  |
Db      3 LNISSL 8


RESULT   4
Q9XNP8
ID   Q9XNP8      PRELIMINARY;      PRT;     8 AA.
AC   Q9XNP8;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   ATP SYNTHASE 6 (FRAGMENT).
GN   ATP6.
OS   Boophilus microplus (Cattle tick).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC   Parasitiformes; Ixodida; Ixodidae; Boophilus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=N;
RX   MEDLINE; 99297341.
RA   Campbell N.J.H., Barker S.C.;
RT   "The novel mitochondrial gene arrangement of the cattle tick,
RT   Boophilus microplus: fivefold tandem repetition of a coding region.";
RL   Mol. Biol. Evol. 16:732-740(1999).
DR   EMBL; AF110616; AAD28386.1; -.
KW   Mitochondrion.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  938 MW;  58BB14404B5735B0 CRC64;


  Query Match             32.6%;  Score 15;  DB 8;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YLN 3
          |:|
Db      4 YIN 6


RESULT   5
P72279
ID   P72279      PRELIMINARY;      PRT;     8 AA.
AC   P72279;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   BIPHENYL DIOXYGENASE (FRAGMENT).
```

```
GN   BPHB.
OS   Rhodococcus globerulus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95255652.
RA   Asturias J.A., Diaz E., Timmis K.N.;
RT   "The evolutionary relationship of biphenyl dioxygenase from gram-
RT   positive Rhodococcus globerulus P6 to multicomponent dioxygenases from
RT   gram-negative bacteria.";
RL   Gene 156:11-18(1995).
DR   EMBL; X80041; CAA56350.1; -.
KW   Dioxygenase.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  989 MW;  EBD2CB1AB6D73406 CRC64;

  Query Match             30.4%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LNVQD 6
              : :||
Db          1 MRLQD 5


RESULT   6
Q50832
ID   Q50832      PRELIMINARY;      PRT;     9 AA.
AC   Q50832;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1996 (TrEMBLrel. 01, Last annotation update)
DE   INTERGENIC AT-RICH DNA SEQUENCE (FRAGMENT).
OS   Methanococcus voltae.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85230552.
RA   Bollschweiler C., Kuehn R., Klein A.;
RT   "Non-repetitive AT-rich sequences are found in intergenic regions of
RT   Methanococcus voltae DNA.";
RL   EMBO J. 4:805-809(1985).
DR   EMBL; X02518; CAA26355.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1087 MW;  99ED005DC404405A CRC64;

  Query Match             30.4%;  Score 14;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          6 DLN 8
              |:|
Db          2 DIN 4


RESULT   7
O98866
ID   O98866      PRELIMINARY;      PRT;     7 AA.
AC   O98866;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CYTOCHROME B/F SUBUNIT IV (FRAGMENT).
OS   Spinacia oleracea (Spinach).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; Caryophyllidae; Caryophyllales;
OC   Chenopodiaceae; Spinacia.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86120353.
RA   Sijben-Mueller G., Hallick R.B., Alt J., Westhoff P., Herrmann R.G.;
RT   "Spinach plastid genes coding for initiation factor IF-1, ribosomal
RT   protein S11 and RNA polymerase alpha-subunit.";
RL   Nucleic Acids Res. 14:1029-1044(1986).
DR   EMBL; X03496; CAA27215.1; -.
KW   Chloroplast.
FT   NON_TER       1      1
SQ   SEQUENCE   7 AA;  907 MW;  644729D77409C420 CRC64;

  Query Match             28.3%;  Score 13;  DB 8;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLN 3
              :||
Db          5 FLN 7


RESULT   8
Q35792
ID   Q35792      PRELIMINARY;      PRT;     8 AA.
AC   Q35792;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   OXI3 LOCUS IN YEAST MITOCHONDRION (STRAIN D273-10B)
DE   (CODES FOR CYTOCHROME OXIDASE SUBUNIT 1).
OS   Saccharomyces cerevisiae (Baker's yeast).
OG   Mitochondrion.
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 81069885.
RA   Bonitz S.G., Coruzzi G., Thalenfeld B.E., Tzagoloff A., Macino G.;
RT   "Assembly of the mitochondrial membrane system. Structure and
RT   nucleotide sequence of the gene coding for subunit 1 of yeast
RT   cytochrme oxidase.";
RL   J. Biol. Chem. 255:11927-11941(1980).
DR   EMBL; V00694; CAA24063.1; -.
KW   Mitochondrion.
SQ   SEQUENCE   8 AA;  1025 MW;  084693345B5AA337 CRC64;

  Query Match             28.3%;  Score 13;  DB 8;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 VQDLN 8
              ::| |
Db          1 LKDYN 5


RESULT   9
Q40659
ID   Q40659      PRELIMINARY;      PRT;     8 AA.
AC   Q40659;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ALPHA-AMYLASE (FRAGMENT).
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91078641.
RA   Kumagai M.H., Shah M., Terashima M., Vrkljan Z., Whitaker J.R.,
RA   Rodriquez R.L.;
RT   "Expression and secretion of rice alpha-amylase by Saccharomyces
```

```
RT   cerevisiae.";
RL   Gene 94:209-216(1990).
DR   EMBL; M62916; AAA33892.1; -.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  948 MW;  EBC694444732D6D6 CRC64;

  Query Match             28.3%;  Score 13;  DB 10;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 VQDLN 8
            :| ||
Db        1 MQVLN 5


RESULT  10
Q48686
ID   Q48686      PRELIMINARY;      PRT;     9 AA.
AC   Q48686;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   PROMOTER 23 DNA FRAGMENT (FRAGMENT).
OS   Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Lactococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88105390.
RA   van der Vossen J.M., der Lelie D., Venema G.;
RT   "Isolation and characterization of Streptococcus cremoris Wg2-specific
RT   promoters.";
RL   Appl. Environ. Microbiol. 53:2452-2457(1987).
DR   EMBL; M24763; AAA74720.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1080 MW;  5AF3A44AA4469443 CRC64;

  Query Match             28.3%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        3 NVQD 6
            |: |
Db        3 NMND 6


RESULT  11
Q14715
ID   Q14715      PRELIMINARY;      PRT;     9 AA.
AC   Q14715;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   KERATIN 14 (FRAGMENT).
GN   KRT14.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92005680.
RA   Coulombe P., Hutton M., Letai A., Hebert A., Paller A., Fuchs E.;
RT   "Point mutations in human keratin 14 genes of epidermolysis bullosa
RT   simplex patients: genetic and functional analyses.";
RL   Cell 66:1301-1311(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95072587.
RA   Yamanishi K., Matsuki M., Konishi K., Yasuno H.;
RT   "A novel mutation of Leu122 to Phe at a highly conserved hydrophobic
RT   residue in the helix initiation motif of keratin 14 in epidermolysis
```

```
RT   bullosa simplex.";
RL   Hum. Mol. Genet. 3:1171-1172(1994).
DR   EMBL; D28807; BAA05967.1; -.
KW   Keratin.
FT   NON_TER       1      1
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1138 MW;  BE300AA449C456D6 CRC64;

  Query Match             28.3%;  Score 13;  DB 4;  Length 9;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 VQDLN 8
            :|: |
Db        2 MQNFN 6


RESULT  12
O95953
ID   O95953      PRELIMINARY;      PRT;     9 AA.
AC   O95953;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   GALACTOCEREBROSIDASE (EC 3.2.1.46) (FRAGMENT).
GN   GALC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Lulli L., Torchiana E., Finocchiaro G.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U77631; AAD15626.1; -.
KW   Hydrolase; Glycosidase.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1069 MW;  374E2AADC2C699C8 CRC64;

  Query Match             28.3%;  Score 13;  DB 4;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 VQDL 7
            | ||
Db        5 VADL 8


RESULT  13
Q9XJN0
ID   Q9XJN0      PRELIMINARY;      PRT;     9 AA.
AC   Q9XJN0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   P10 (FRAGMENT).
OS   bacteriophage phi-10.
OC   Viruses; dsRNA viruses; Cystoviridae; Cystovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mindich L., Qiao X., Qiao J., Romantschuk M., Hoogstraten D.;
RT   "Isolation of additional bacteriophages with genomes of segmented
RT   double-stranded RNA.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF125675; AAD22555.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1058 MW;  880E376AA720544A CRC64;

  Query Match             28.3%;  Score 13;  DB 9;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
```

```
  Matches     2;  Conservative     1;  Mismatches     1;  Indels     0;  Gaps     0;

Qy       3 NVQD 6
           |: |
Db       3 NILD 6


RESULT  14
Q9RQ57
ID   Q9RQ57      PRELIMINARY;      PRT;     8 AA.
AC   Q9RQ57;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   NIFS PROTEIN HOMOLOG (FRAGMENT).
GN   NIFS.
OS   Buchnera aphidicola.
OC   Bacteria; Proteobacteria; gamma subdivision; Buchnera.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 20022990.
RA   Clark M.A., Moran N.A., Baumann P.;
RT   "Sequence evolution in bacterial endosymbionts having extreme base
RT   compositions.";
RL   Mol. Biol. Evol. 16:1586-1598(1999).
DR   EMBL; AF130812; AAF13797.1; -.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  980 MW;  F3A73B504771A336 CRC64;


  Query Match             26.1%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches     2;  Conservative     1;  Mismatches     0;  Indels     0;  Gaps     0;

Qy       1 YLN 3
           ||:
Db       6 YLD 8


RESULT  15
Q9RQ49
ID   Q9RQ49      PRELIMINARY;      PRT;     8 AA.
AC   Q9RQ49;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   NIFS PROTEIN HOMOLOG (FRAGMENT).
GN   NIFS.
OS   Buchnera aphidicola.
OC   Bacteria; Proteobacteria; gamma subdivision; Buchnera.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 20022990.
RA   Clark M.A., Moran N.A., Baumann P.;
RT   "Sequence evolution in bacterial endosymbionts having extreme base
RT   compositions.";
RL   Mol. Biol. Evol. 16:1586-1598(1999).
DR   EMBL; AF130814; AAF13805.1; -.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  992 MW;  F3A73B5047672336 CRC64;


  Query Match             26.1%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches     2;  Conservative     1;  Mismatches     0;  Indels     0;  Gaps     0;

Qy       1 YLN 3
           ||:
Db       6 YLD 8


RESULT  16
P87225
ID   P87225      PRELIMINARY;      PRT;     8 AA.
AC   P87225;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GIN11 PROTEIN (FRAGMENT).
GN   GIN11.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wedler H., Wedler E., Scharfe M., Wambutt R.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z73169; CAA97518.2; -.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  1019 MW;  4E21A9C449D5B73B CRC64;


  Query Match             26.1%;  Score 12;  DB 3;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches     2;  Conservative     1;  Mismatches     0;  Indels     0;  Gaps     0;

Qy       1 YLN 3
           ||:
Db       1 YLS 3


RESULT  17
Q15889
ID   Q15889      PRELIMINARY;      PRT;     8 AA.
AC   Q15889;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   (CLONE XP15H8B) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   Lee C.C., Yazdani A., Wehnert M., Bailey J., Couch L., Xiong M.,
RA   Coolbaugh M.I., Chinault C.A., Baldini A., Lindsay E.A., Zhao Z.Y.,
RA   Caskey C.T.H.;
RL   Hum. Mol. Genet. 0:0-0(0).
DR   EMBL; L32070; AAA73879.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  865 MW;  0474472325A761E7 CRC64;


  Query Match             26.1%;  Score 12;  DB 4;  Length 8;
  Best Local Similarity   42.9%;  Pred. No. 3e+05;
  Matches     3;  Conservative     1;  Mismatches     3;  Indels     0;  Gaps     0;

Qy       2 LNVQDLN 8
           |:   ||
Db       1 LHPSKLN 7


RESULT  18
Q9UMH9
ID   Q9UMH9      PRELIMINARY;      PRT;     8 AA.
AC   Q9UMH9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RHCE PROTEIN (FRAGMENT).
```

```
GN   RHCE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RA   Matassi G., Cherif-Zahar B., Mouro I., Cartron J.P.;
RT   "Characterization of the recombination hot spot involved in the
RT   genomic rearrangement leading to the hybrid D-CE-D gene in the DVI
RT   phenotype.";
RL   Am. J. Hum. Genet. 60:808-817(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE; 90349591.
RA   Cherif-Zahar B., Bloy C., Le Van Kim C., Blanchard D., Bailly P.,
RA   Hermand P., Salmon C., Cartron J.-P., Colin Y.;
RT   "Molecular cloning and protein structure of a human blood group Rh
RT   polypeptide.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:6243-6247(1990).
DR   EMBL; Z97030; CAB09726.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  1049 MW;  C007244691FB5AB1 CRC64;

  Query Match             26.1%;  Score 12;  DB 4;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        7 LNL 9
            :||
Db        5 MNL 7

RESULT  19
Q9UL56
ID   Q9UL56      PRELIMINARY;      PRT;     8 AA.
AC   Q9UL56;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MUTANT NADH-CYTOCHROME B5 REDUCTASE (EC 1.6.2.2) (FRAGMENT).
GN   DIA1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fukumaki Y., Higasa K.;
RT   "Two novel mutations in Thai patients with hereditary
RT   methemoglobinemia types I and II: a subtle amino acid change causes
RT   instability of NADH-cytochrome b5 reductase.";
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF061830; AAF06818.1; -.
KW   Oxidoreductase.
FT   NON_TER       1      1
FT   VARIANT       9      9       -> R.
SQ   SEQUENCE   8 AA;  888 MW;  76C5B73B5051F6D8 CRC64;

  Query Match             26.1%;  Score 12;  DB 4;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLN 3
            ||:
Db        5 YLS 7

RESULT  20
Q57328
ID   Q57328      PRELIMINARY;      PRT;     9 AA.
AC   Q57328;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (STRAIN 132) (FRAGMENT).
GN   EXEF'.
OS   Aeromonas sobria.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=132, FROM DR M. ALTWEGG UNIV. ZURICH CULTRE COLLECTION;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89466; CAA61647.1; -.
DR   EMBL; X89465; CAA61645.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1077 MW;  79E852C69444472B CRC64;

  Query Match             26.1%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 DLN 8
            :||
Db        2 ELN 4

RESULT  21
Q44001
ID   Q44001      PRELIMINARY;      PRT;     9 AA.
AC   Q44001;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (STRAIN CDC 9179-79) (FRAGMENT).
GN   EXEF'.
OS   Aeromonas eucrenophila.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 9179-79;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89461; CAA61637.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1059 MW;  785852C69444472B CRC64;

  Query Match             26.1%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 DLN 8
            :||
Db        2 ELN 4

RESULT  22
Q44377
ID   Q44377      PRELIMINARY;      PRT;     9 AA.
AC   Q44377;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
```

```
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (STRAIN ATCC 49659) (FRAGMENT).
GN   EXEF'.
OS   Aeromonas trota.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 49659;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89468; CAA61651.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1059 MW;  785852C69444472B CRC64;


  Query Match             26.1%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DLN 8
           :||
Db       2 ELN 4


RESULT  23
Q44468
ID   Q44468      PRELIMINARY;      PRT;     9 AA.
AC   Q44468;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (STRAIN CDC 1306-83) (FRAGMENT).
GN   EXEF'.
OS   Aeromonas veronii.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1306-83;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89457; CAA61629.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1059 MW;  785852C69444472B CRC64;


  Query Match             26.1%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DLN 8
           :||
Db       2 ELN 4


RESULT  24
Q43928
ID   Q43928      PRELIMINARY;      PRT;     9 AA.
AC   Q43928; Q43918; Q43920; Q43921;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (FRAGMENT).
GN   EXEF'.
OS   Aeromonas caviae.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
```

```
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VARIOUS STRAINS;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89464; CAA61643.1; -.
DR   EMBL; X89462; CAA61639.1; -.
DR   EMBL; X89460; CAA61635.1; -.
DR   EMBL; X89463; CAA61641.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1059 MW;  785852C69444472B CRC64;


  Query Match             26.1%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DLN 8
           :||
Db       2 ELN 4


RESULT  25
Q9UQW0
ID   Q9UQW0      PRELIMINARY;      PRT;     9 AA.
AC   Q9UQW0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   PROLACTIN PRECURSOR (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84182507.
RA   Truong A.T., Duez C., Belayew A., Renard A., Pictet R., Bell G.I.,
RA   Martial J.A.;
RT   "Isolation and characterization of the human prolactin gene.";
RL   EMBO J. 3:429-437(1984).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93076813.
RA   Peers B., Nalda A.M., Monget P., Voz M.L., Belayew A., Martial J.A.;
RT   "Binding of a 100-kDa ubiquitous factor to the human prolactin
RT   promoter is required for its basal and hormone-regulated activity.";
RL   Eur. J. Biochem. 210:53-58(1992).
DR   EMBL; X00368; CAA25108.1; -.
KW   Signal.
FT   SIGNAL        1      8       POTENTIAL.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1060 MW;  0A1A6775B8733054 CRC64;


  Query Match             26.1%;  Score 12;  DB 4;  Length 9;
  Best Local Similarity   25.0%;  Pred. No. 3e+05;
  Matches    1;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LNVQ 5
           :|::
Db       1 MNIK 4


RESULT  26
Q9UQA3
ID   Q9UQA3      PRELIMINARY;      PRT;     9 AA.
AC   Q9UQA3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
```

```
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   P27KIP1 PROTEIN (FRAGMENT).
GN   P27KIP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ito E., Iwahashi Y., Yanagisawa Y., Suzuki Y., Sugano S., Yuasa Y.,
RA   Maruyama K.;
RT   "Two short sequences have positive effects on the human p27(Kip1) gene
RT   transcription.";
RL   Gene 228:93-100(1999).
DR   EMBL; AB005590; BAA76715.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  963 MW;  984D55A2D412C455 CRC64;


  Query Match             26.1%;  Score 12;  DB 4;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 NVQDLN 8
           ||:  |
Db       3 NVRVSN 8


RESULT  27
O96417
ID   O96417      PRELIMINARY;      PRT;     9 AA.
AC   O96417;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   SXL E1 FORM (FRAGMENT).
GN   SXL.
OS   Drosophila virilis (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98337843.
RA   Erickson J.W., Cline T.W.;
RT   "Key aspects of the primary sex determination mechanism are conserved
RT   across the genus Drosophila.";
RL   Development 125:3259-3268(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Bell M., Cline T.W.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF046045; AAC97605.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1089 MW;  9A8BD1AAA9C449CA CRC64;


  Query Match             26.1%;  Score 12;  DB 5;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       6 DLN 8
             | |
Db       2 DFN 4


RESULT  28
Q88612
ID   Q88612      PRELIMINARY;      PRT;     9 AA.
AC   Q88612;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   STRUCTURAL POLYPROTEIN (FRAGMENT).
OS   Middelburg virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC   Alphavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83039346.
RA   Ou J.H., Rice C.M., Dalgarno L., Strauss E.G., Strauss J.H.;
RT   "Sequence studies of several alphavirus genomic RNAs in the region
RT   containing the start of the subgenomic RNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 79:5235-5239(1982).
DR   EMBL; J02246; AAA96655.1; -.
KW   Polyprotein.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1114 MW;  00E8B6C1B7604B54 CRC64;

  Query Match             26.1%;  Score 12;  DB 12;  Length 9;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLNVQ 5
             |:  |
Db       3 YIPTQ 7


RESULT  29
Q89491
ID   Q89491      PRELIMINARY;      PRT;     9 AA.
AC   Q89491;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 1.1 KDA PROTEIN.
OS   Murine minute virus (Murine parvovirus).
OC   Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LYMPHOTROPIC VARIANT;
RX   MEDLINE; 86115415.
RA   Astell C.R., Gardiner E.M., Tattersall P.;
RT   "DNA sequence of the lymphotropic variant of minute virus of mice,
RT   MVM(i), and comparison with the DNA sequence of the fibrotropic
RT   prototype strain.";
RL   J. Virol. 570:656-669(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MVM(P);
RX   MEDLINE; 83143341.
RA   Astell C.R., Thomson M., Merchlinsky M., Ward D.C.;
RT   "The complete DNA sequence of minute virus of mice, an autonomous
RT   parvovirus.";
RL   Nucleic Acids Res. 11:999-1018(1983).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MVM(P);
RX   MEDLINE; 86115415.
RA   Astell C.R., Gardiner E.M., Tattersall P.;
RT   "DNA sequence of the lymphotropic variant of minute virus of mice,
RT   MVM(i), and comparison with the DNA sequence of the fibrotropic
RT   prototype strain.";
RL   J. Virol. 57:656-669(1986).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MVM(P);
RX   MEDLINE; 87061199.
RA   Morgan W.R., Ward D.C.;
RT   "Three splicing patterns are used to excise the small intron common to
RT   all minute virus of mice RNAs.";
RL   J. Virol. 60:1170-1174(1986).
DR   EMBL; M12032; AAA69570.1; -.
DR   EMBL; J02275; AAA67112.1; -.
DR   EMBL; V01115; CAA24311.1; -.
KW   Hypothetical protein.
```

```
SQ   SEQUENCE   9 AA;  1061 MW;  C3FD405863637862 CRC64;

  Query Match             26.1%;  Score 12;  DB 12;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LNV 4
            :||
Db        7 INV 9

RESULT  30
Q9YQ10
ID   Q9YQ10      PRELIMINARY;      PRT;     7 AA.
AC   Q9YQ10;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HYPOTHETICAL FUSION PROTEIN.
OS   porcine transmissible gastroenteritis virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99099045.
RA   Izeta A., Smerdou C., Alonso S., Penzes Z., Mendez A., Plana-Duran J.,
RA   Enjuanes L.;
RT   "Replication and packaging of transmissible gastroenteritis
RT   coronavirus-derived synthetic minigenomes.";
RL   J. Virol. 73:1535-1545(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95159435.
RA   Eleouet J., Rasschaert D., Lambert P., Levy L., Vende P., Laude H.;
RT   "Complete sequence (20 kilobases) of the polyprotein-encoding gene 1
RT   of transmissible gastroenteritis virus.";
RL   Virology 206:817-822(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88078100.
RA   Rasschaert D., Gelfi J., Laude H.;
RT   "Enteric coronavirus TGEV: partial sequence of the genomic RNA, its
RT   organization and expression.";
RL   Biochimie 69:591-600(1987).
DR   EMBL; AJ011482; CAA09625.1; -.
SQ   SEQUENCE   7 AA;  927 MW;  69D6D7273B5726F0 CRC64;

  Query Match             23.9%;  Score 11;  DB 12;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YL 2
            ||
Db        3 YL 4

RESULT  31
Q51594
ID   Q51594      PRELIMINARY;      PRT;     8 AA.
AC   Q51594;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   COPB PROTEIN (FRAGMENT).
OS   Escherichia coli.
OG   Plasmid IncFI ColV2-K94.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86223772.
RA   Weber P.C., Palchaudhuri S.;
RT   "Incompatibility repressor in a RepA-like replicon of the IncFI
RT   plasmid ColV2-K94.";
RL   J. Bacteriol. 166:1106-1112(1986).
DR   EMBL; M13472; AAA23194.1; -.
KW   Plasmid.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  998 MW;  0F37205AA73416D7 CRC64;

  Query Match             23.9%;  Score 11;  DB 2;  Length 8;
  Best Local Similarity   33.3%;  Pred. No. 3e+05;
  Matches    2;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 VQDLNL 9
            :| |::
Db        1 LQRLDI 6

RESULT  32
Q9ZEZ9
ID   Q9ZEZ9      PRELIMINARY;      PRT;     8 AA.
AC   Q9ZEZ9;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   2-ISOPROPYLMALATE SYNTHASE (EC 4.1.3.12) (FRAGMENT).
GN   LEUA.
OS   Buchnera aphidicola.
OG   Plasmid pBRcl.
OC   Bacteria; Proteobacteria; gamma subdivision; Buchnera.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99028904.
RA   Silva F.J., Van Ham R.C.H.J., Sabater B., Latorre A.;
RT   "Structure and evolution of the leucine plasmids carried by the
RT   endosymbiont (Buchnera aphidicola) from aphids of the family
RT   Aphididae.";
RL   FEMS Microbiol. Lett. 168:43-49(1998).
DR   EMBL; AJ006874; CAA07290.1; -.
KW   Lyase; Plasmid.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  917 MW;  FE205042D6C5B446 CRC64;

  Query Match             23.9%;  Score 11;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        2 LNVQ 5
            :| |
Db        1 MNSQ 4

RESULT  33
Q9UDZ4
ID   Q9UDZ4      PRELIMINARY;      PRT;     8 AA.
AC   Q9UDZ4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RHD PROTEIN (FRAGMENT).
GN   RHD.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RA   Matassi G., Cherif-Zahar B., Mouro I., Cartron J.P.;
RT   "Characterization of the recombination hot spot involved in the
RT   genomic rearrangement leading to the hybrid D-CE-D gene in the DVI
RT   phenotype.";
```

```
RL   Am. J. Hum. Genet. 60:808-817(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE; 93066356.
RA   Le Van Kim C., Mouro I., Cherif-Zahar B., Raynal V., Cherrier C.,
RA   Cartron J.P., Colin Y.;
RT   "Molecular cloning and primary structure of the human blood group RhD
RT   polypeptide.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:10925-10929(1992).
DR   EMBL; Z97031; CAB09727.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  1042 MW;  D296944691FB5AB1 CRC64;

  Query Match             23.9%;  Score 11;  DB 4;  Length 8;
  Best Local Similarity   25.0%;  Pred. No. 3e+05;
  Matches    1;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLNV 4
           ::|:
Db       4 HMNM 7


RESULT  34
Q34909
ID   Q34909      PRELIMINARY;      PRT;     8 AA.
AC   Q34909;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CYTOCHROME B (FRAGMENT).
OS   Locusta migratoria (Migratory locust).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC   Acridomorpha; Acridoidea; Acrididae; Oedipodinae; Locusta.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88223478.
RA   McCracken A., Uhlenbusch I., Gellissen G.;
RT   "Structure of the cloned Locusta migratoria mitochondrial genome:
RT   restriction mapping and sequence of its ND-1 (URF-1) gene.";
RL   Curr. Genet. 11:625-630(1987).
DR   EMBL; X05286; CAA28905.1; -.
KW   Mitochondrion.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  1019 MW;  F8E33723304B45B6 CRC64;

  Query Match             23.9%;  Score 11;  DB 8;  Length 8;
  Best Local Similarity   20.0%;  Pred. No. 3e+05;
  Matches    1;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLNVQ 5
           |: ::
Db       3 YIKLK 7


RESULT  35
Q9T2W0
ID   Q9T2W0      PRELIMINARY;      PRT;     8 AA.
AC   Q9T2W0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CBS1 PRECURSOR=PROTEOLYTIC PROCESSING SITE.
OS   Saccharomyces cerevisiae (Baker's yeast).
OG   Mitochondrion.
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92035078.
RA   Korte A., Michaelis U., Lottspeich F., Rodel G.;
RT   "Over-expression, purification and determination of the proteolytic
RT   processing site of the yeast mitochondrial CBS1 protein.";
RL   Curr. Genet. 20:87-90(1991).
SQ   SEQUENCE   8 AA;  913 MW;  72D1A44041B40047 CRC64;

  Query Match             23.9%;  Score 11;  DB 8;  Length 8;
  Best Local Similarity   20.0%;  Pred. No. 3e+05;
  Matches    1;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 VQDLN 8
           :: :|
Db       2 IRTIN 6


Search completed: December 16, 2000, 04:22:13
Job time: 4607 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 03:35:13 ; Search time 89.11 Seconds
                                              (without alignments)
                                              6.409 Million cell updates/sec

Title:          US-09-529-121-9
Perfect score:  46
Sequence:       1 YLNVQDLNL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters:      793

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       PIR_65:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                   %
Result           Query
  No.   Score    Match Length DB  ID                    Description
-----------------------------------------------------------------------
    1      21    45.7      9  1  AKLQIM                locustamyoinhibiti
    2      21    45.7      9  2  A57444                neuropeptide Grb-A
    3      17    37.0      9  2  B57444                neuropeptide Grb-A
    4      16    34.8      7  2  PC2370                probable H+-transp
    5      15    32.6      9  2  PL0139                carbon-monoxide de
    6      14    30.4      7  2  PN0649                alpha-dextrin endo
    7      13    28.3      6  2  B35640                cerebellar degener
    8      12    26.1      7  2  B39040                calsequestrin, fas
    9      12    26.1      7  2  S58797                serine/threonine-s
   10      12    26.1      8  2  PQ0012                cholecystokinin -
   11      12    26.1      8  2  A43001                cholecystokinin -
   12      12    26.1      8  2  T13818                cytochrome oxidase
   13      12    26.1      8  2  PC4131                hypothetical prote
   14      12    26.1      8  2  S22428                chitin-binding pro
   15      12    26.1      8  2  S21663                neuropeptide - flo
   16      12    26.1      9  2  D48186                ATPase R1 subunit
   17      12    26.1      9  2  C57444                neuropeptide Grb-A
   18      12    26.1      9  2  A37027                macrophage chemota
   19      12    26.1      9  2  B30572                T-cell receptor be
   20      11    23.9      5  1  HOROHA                proctolin - Americ
   21      11    23.9      5  2  B31836                20K protein - Rick
   22      11    23.9      5  2  A60411                proctolin - Atlant
   23      11    23.9      5  2  PT0278                Ig heavy chain CRD
   24      11    23.9      6  2  B44510                hypothetical prote
   25      11    23.9      6  2  JH0784                neuropeptide TE-6
   26      11    23.9      7  2  S78024                ribosomal protein
   27      11    23.9      7  2  A11483                aspartate transami
   28      11    23.9      8  2  PQ0726                unidentified 4.5/4
   29      11    23.9      8  2  PT0311                Ig heavy chain CRD
```

```
   30      11    23.9      8  2  A42689                major postsynaptic
   31      11    23.9      9  2  A44873                caldesmon - rabbit
   32      11    23.9      9  2  A61230                calsequestrin, car
   33      11    23.9      9  2  A60108                exotoxin A - Strep
   34      11    23.9      9  2  A41978                calliFMRFamide 1 -
   35      11    23.9      9  2  D44787                calliFMRFamide 13
   36      11    23.9      9  2  B41978                calliFMRFamide 2 -
   37      11    23.9      9  2  C41978                calliFMRFamide 3 -
   38      11    23.9      9  2  D41978                calliFMRFamide 4 -
   39      11    23.9      9  2  E41978                calliFMRFamide 5 -
   40      11    23.9      9  2  F41978                calliFMRFamide 6 -
   41      11    23.9      9  2  G41978                calliFMRFamide 7 -
   42      11    23.9      9  2  PT0299                Ig heavy chain CRD
   43      11    23.9      9  2  B39504                octamer-binding pr
   44      11    23.9      9  2  S56004                glucan 1,3-beta-gl
   45      11    23.9      9  4  I73804                hypothetical E2 pr
   46      10    21.7      4  2  A37832                phenol 2-monooxyge
   47      10    21.7      4  2  A26209                protein-glutamine
   48      10    21.7      6  2  PC4392                whey glycoprotein
   49      10    21.7      7  2  S71867                glutathione transf
   50      10    21.7      7  2  S20446                elastase - Pseudom
   51      10    21.7      7  2  I40504                hypothetical prote
   52      10    21.7      7  2  A34026                acetylcholinestera
   53      10    21.7      7  2  S21230                dermorphin (Trp-4,
   54      10    21.7      7  2  S68004                hucolin, 75K chain
   55      10    21.7      7  2  C56793                platelet glycoprot
   56      10    21.7      7  2  S45648                Na+-transporting A
   57      10    21.7      8  2  A61348                red pigment-concen
   58      10    21.7      8  2  S15422                adipokinetic hormo
   59      10    21.7      8  2  A43976                hypertrehalosemic
   60      10    21.7      8  2  B43976                hypertrehalosemic
   61      10    21.7      8  2  A58641                adipokinetic hormo
   62      10    21.7      8  2  S68802                nitrate reductase
   63      10    21.7      8  2  A14683                aspartate transami
   64      10    21.7      8  2  I49404                prealbumin - weste
   65      10    21.7      8  2  S65381                cytochrome-c oxida
   66      10    21.7      8  2  I57532                gene TnIslow prote
   67      10    21.7      8  2  S69165                ferredoxin a2 - Ja
   68      10    21.7      9  2  A91466                oxytocin - hippopo
   69      10    21.7      9  2  A92774                oxytocin - spotted
   70      10    21.7      9  2  A93147                oxytocin - finback
   71      10    21.7      9  2  A93408                oxytocin - Austral
   72      10    21.7      9  2  B90667                oxytocin - rabbit
   73      10    21.7      9  2  A61364                isotocin - common
   74      10    21.7      9  2  C24180                fibrinogen beta ch
   75      10    21.7      9  2  C41170                photosystem II pro

                              ALIGNMENTS

RESULT   1
AKLQIM
locustamyoinhibiting peptide - migratory locust
C;Species: Locusta migratoria (migratory locust)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 20-Mar-1998
C;Accession: A60065
R;Schoofs, L.; Holman, G.M.; Hayes, T.K.; Nachman, R.J.; De Loof, A.
Regul. Pept. 36, 111-119, 1991
A;Title: Isolation, identification and synthesis of locustamyoinhibiting peptide (LOM
A;Reference number: A60065; MUID:92179466
A;Accession: A60065
A;Molecule type: protein
A;Residues: 1-9 <SCH>
C;Comment: This peptide hormone suppresses spontaneous contractions of the hindgut an
C;Superfamily: locustamyoinhibiting peptide
C;Keywords: amidated carboxyl end; hormone
F;9/Modified site: amidated carboxyl end (Trp) #status experimental


  Query Match             45.7%;  Score 21;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       5 QDLN 8
           ||||
Db       3 QDLN 6


RESULT   2
A57444
neuropeptide Grb-AST B1 - two-spotted cricket
C;Species: Gryllus bimaculatus (two-spotted cricket)
C;Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 26-Jan-1996
C;Accession: A57444
R;Lorenz, M.W.; Kellner, R.; Hoffmann, K.H.
J. Biol. Chem. 270, 21103-21108, 1995
A;Title: A family of neuropeptides that inhibit juvenile hormone biosynthesis in the cri
A;Reference number: A57444; MUID:95403341
A;Accession: A57444
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <LOR>


  Query Match             45.7%;  Score 21;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 QDLN 8
           ||||
Db       3 QDLN 6


RESULT   3
B57444
neuropeptide Grb-AST B2 - two-spotted cricket
C;Species: Gryllus bimaculatus (two-spotted cricket)
C;Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 26-Jan-1996
C;Accession: B57444
R;Lorenz, M.W.; Kellner, R.; Hoffmann, K.H.
J. Biol. Chem. 270, 21103-21108, 1995
A;Title: A family of neuropeptides that inhibit juvenile hormone biosynthesis in the cri
A;Reference number: A57444; MUID:95403341
A;Accession: B57444
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <LOR>


  Query Match             37.0%;  Score 17;  DB 2;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 QDLN 8
           :|||
Db       3 RDLN 6


RESULT   4
PC2370
probable H+-transporting ATP synthase (EC 3.6.1.34) alpha chain [similarity] - Bacillus
N;Alternate names: unidentified 78K protein
C;Species: Bacillus cereus
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: PC2370
R;Matsuno, K.; Miyamoto, T.; Yamaguchi, K.; Sayed, M.A.; Kajiwara, T.; Hatano, S.
Biosci. Biotechnol. Biochem. 59, 231-235, 1995
A;Title: Identification of DNA-binding proteins changed after induction of sporulation i
A;Reference number: PC2369; MUID:95218265
A;Accession: PC2370
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-7 <MAS>
C;Keywords: ATP biosynthesis; hydrolase
```

```
  Query Match             34.8%;  Score 16;  DB 2;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DLN 8
           |||
Db       2 DLN 4


RESULT   5
PL0139
carbon-monoxide dehydrogenase (EC 1.2.99.2) large chain - Pseudomonas carboxydoflava
C;Species: Pseudomonas carboxydoflava
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 28-Apr-1993
C;Accession: PL0139
R;Kraut, M.; Hugendieck, I.; Herwig, S.; Meyer, O.
Arch. Microbiol. 152, 335-341, 1989
A;Title: Homology and distribution of CO dehydrogenase structural genes in carboxydot
A;Reference number: PL0138; MUID:90055678
A;Accession: PL0139
A;Molecule type: protein
A;Residues: 1-9 <KRA>
A;Note: 2-Met is also found
C;Comment: Cardon-monoxide dehydrogenase consists of three polypeptide chains: large,
C;Keywords: oxidoreductase


  Query Match             32.6%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 VQD 6
           |||
Db       5 VQD 7


RESULT   6
PN0649
alpha-dextrin endo-1,6-alpha-glucosidase (EC 3.2.1.41) - Bacillus sp. (strain S-1) (f
C;Species: Bacillus sp.
C;Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 07-May-1999
C;Accession: PN0649
R;Kim, C.H.; Choi, H.I.; Lee, D.S.
Biosci. Biotechnol. Biochem. 57, 1632-1637, 1993
A;Title: Purification and biochemical properties of an alkaline pullulanase from alka
A;Reference number: PN0649; MUID:94080025
A;Accession: PN0649
A;Molecule type: protein
A;Residues: 1-7 <KIM>
C;Comment: This enzyme is used togther with glucoamylase to improve the efficiency of
ntent in high maltose syrups.
C;Keywords: glycosidase; hydrolase; polysaccharide degradation


  Query Match             30.4%;  Score 14;  DB 2;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLNV 4
           :||:
Db       1 FLNM 4


RESULT   7
B35640
cerebellar degeneration-related protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 28-Sep-1990 #sequence_revision 28-Sep-1990 #text_change 24-Jun-1993
C;Accession: B35640
R;Chen, Y.T.; Rettig, W.J.; Yenamandra, A.K.; Kozak, C.A.; Chaganti, R.S.K.; Posner,
Proc. Natl. Acad. Sci. U.S.A. 87, 3077-3081, 1990
A;Title: Cerebellar degeneration-related antigen: a highly conserved neuroectodermal
```

```
A;Reference number: A35640; MUID:90222173
A;Accession: B35640
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-6 <CHE>


  Query Match             28.3%;  Score 13;  DB 2;  Length 6;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         5 QDLN 8
               :||:
Db         3 EDLD 6


RESULT   8
B39040
calsequestrin, fast skeletal muscle - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 23-Aug-1991 #sequence_revision 23-Aug-1991 #text_change 23-Feb-1997
C;Accession: B39040
R;Cala, S.E.; Jones, L.R.
J. Biol. Chem. 266, 391-398, 1991
A;Title: Phosphorylation of cardiac and skeletal muscle calsequestrin isoforms by casei
A;Reference number: A39040; MUID:91093153
A;Accession: B39040
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-7 <CAL>
C;Keywords: phosphoprotein; skeletal muscle


  Query Match             26.1%;  Score 12;  DB 2;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         5 QDL 7
               :||
Db         4 EDL 6


RESULT   9
S58797
serine/threonine-specific protein kinase c-mos - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 23-Feb-1997
C;Accession: S58797
R;Nagao, Y.
Biochim. Biophys. Acta 1245, 130-143, 1995
A;Title: Expression of c-mos protein in cultured rat spermatogenic cells and evidence t
A;Reference number: S58797; MUID:95383384
A;Accession: S58797
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-7 <NAG>
C;Genetics:
A;Gene: c-mos
C;Keywords: phosphotransferase; protein kinase


  Query Match             26.1%;  Score 12;  DB 2;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         3 NVQ 5
               |:|
Db         5 NLQ 7


RESULT  10
PQ0012
cholecystokinin - southeastern quoll
N;Alternate names: CCK
C;Species: Dasyurus viverrinus (southeastern quoll)
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 13-Sep-1996
C;Accession: PQ0012
R;Fan, Z.W.; Eng, J.; Shaw, G.; Yalow, R.S.
Peptides 9, 429-431, 1988
A;Title: Cholecystokinin octapeptide purified from brains of Australian marsupials.
A;Reference number: PQ0012; MUID:88234141
A;Accession: PQ0012
A;Molecule type: protein
A;Residues: 1-8 <FAN>
C;Superfamily: gastrin
C;Keywords: amidated carboxyl end; hormone; neuropeptide; sulfoprotein
F;2/Binding site: sulfate (Tyr) (covalent) #status predicted
F;8/Modified site: amidated carboxyl end (Phe) #status predicted


  Query Match             26.1%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   33.3%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         1 YLNVQD 6
             |:   |
Db         2 YMGWMD 7


RESULT  11
A43001
cholecystokinin - tammar wallaby
N;Alternate names: CCK
C;Species: Macropus eugenii (tammar wallaby)
C;Date: 30-Oct-1992 #sequence_revision 30-Oct-1992 #text_change 13-Sep-1996
C;Accession: A43001; PQ0012
R;Fan, Z.W.; Eng, J.; Shaw, G.; Yalow, R.S.
Peptides 9, 429-431, 1988
A;Title: Cholecystokinin octapeptide purified from brains of Australian marsupials.
A;Reference number: PQ0012; MUID:88234141
A;Accession: A43001
A;Molecule type: protein
A;Residues: 1-8 <FAN>
C;Superfamily: gastrin
C;Keywords: amidated carboxyl end; hormone; neuropeptide; sulfoprotein
F;2/Binding site: sulfate (Tyr) (covalent) #status predicted
F;8/Modified site: amidated carboxyl end (Phe) #status predicted


  Query Match             26.1%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   33.3%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         1 YLNVQD 6
             |:   |
Db         2 YMGWMD 7


RESULT  12
T13818
cytochrome oxidase subunit I - Atlantic hagfish mitochondrion (fragment)
C;Species: mitochondrion Myxine glutinosa (Atlantic hagfish)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C;Accession: T13818
R;Delarbre, C.; Barriel, V.; Tillier, S.; Janvier, P.; Gachelin, G.
Mol. Biol. Evol. 14, 807-813, 1997
A;Title: The main features of the craniate mitochondrial DNA between the ND1 and the
A;Reference number: Z17775; MUID:97398704
A;Accession: T13818
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-8 <DEL>
A;Cross-references: EMBL:Y09527; NID:e1011465; PID:e329906; PIDN:CAA70718.1
C;Genetics:
A;Genome: mitochondrion
A;Note: COI
```

```
C;Keywords: mitochondrion

  Query Match             26.1%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLN 3
              ||:
Db          2 YLS 4


RESULT  13
PC4131
hypothetical protein 8 [imported] - Pseudomonas aeruginosa (fragment)
C;Species: Pseudomonas aeruginosa
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: PC4131
R;Kawasaki, S.; Arai, H.; Igarashi, Y.; Kodama, T.
Gene 167, 87-91, 1995
A;Title: Sequencing and characterization of the downstream region of the genes encoding
y for biosynthesis of heme d1.
A;Reference number: JC4552; MUID:96144254
A;Accession: PC4131
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-8 <KAW>
A;Cross-references: DDBJ:D50473; NID:g1217594
A;Note: this ORF is not annotated in GenBank entry PSENIRC, release 113.0


  Query Match             26.1%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 VQDLN 8
              : ||:
Db          1 MDDLS 5


RESULT  14
S22428
chitin-binding protein - potato (fragment)
C;Species: Solanum tuberosum (potato)
C;Date: 19-Mar-1997 #sequence_revision 10-Oct-1997 #text_change 05-Dec-1998
C;Accession: S22428
R;Millar, D.J.; Allen, A.K.; Smith, C.G.; Sidebottom, C.; Slabas, A.R.; Bolwell, G.P.
Biochem. J. 283, 813-821, 1992
A;Title: Chitin-binding proteins in potato (Solanum tuberosum L.) tuber. Characterizatic
A;Reference number: S21288; MUID:92272683
A;Accession: S22428
A;Molecule type: protein
A;Residues: 1-8 <MIL>
C;Function:
A;Description: may be involved in plant defence
C;Keywords: glycoprotein; hydroxyproline


  Query Match             26.1%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLN 3
              | |
Db          6 YAN 8


RESULT  15
S21663
neuropeptide - flower beetle (Pachnoda marginata)
C;Species: Pachnoda marginata
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
C;Accession: S21663
```

```
R;Gaede, G.; Lopata, A.; Kellner, R.; Rinehart, K.L.
Biol. Chem. Hoppe-Seyler 373, 133-142, 1992
A;Title: Primary structures of neuropeptides isolated from the corpora cardiaca of va
ectrometry.
A;Reference number: S21663; MUID:92265187
A;Accession: S21663
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-8 <GAE>

  Query Match             26.1%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          6 DLN 8
              :||
Db          1 ELN 3


RESULT  16
D48186
ATPase R1 subunit - wood tobacco (fragment)
C;Species: Nicotiana sylvestris (wood tobacco)
C;Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 23-Feb-1997
C;Accession: D48186
R;De Paepe, R.; Forchioni, A.; Chetrit, P.; Vedel, F.
Proc. Natl. Acad. Sci. U.S.A. 90, 5934-5938, 1993
A;Title: Specific mitochondrial proteins in pollen: presence of an additional ATP syn
A;Reference number: A48186; MUID:93317598
A;Accession: D48186
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <DE1>
A;Experimental source: pollen
A;Note: sequence extracted from NCBI backbone (NCBIP:134871)


  Query Match             26.1%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 VQDL 7
              | ||
Db          2 VVDL 5


RESULT  17
C57444
neuropeptide Grb-AST B3 - two-spotted cricket
C;Species: Gryllus bimaculatus (two-spotted cricket)
C;Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 26-Jan-1996
C;Accession: C57444
R;Lorenz, M.W.; Kellner, R.; Hoffmann, K.H.
J. Biol. Chem. 270, 21103-21108, 1995
A;Title: A family of neuropeptides that inhibit juvenile hormone biosynthesis in the
A;Reference number: A57444; MUID:95403341
A;Accession: C57444
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <LOR>


  Query Match             26.1%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 QDLN 8
              :||:
Db          3 RDLS 6


RESULT  18
```

```
A37027
macrophage chemotactic factor - human (fragment)
N;Alternate names: T-cell hybridoma D6-18 protein
C;Species: Homo sapiens (man)
C;Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 23-Feb-1997
C;Accession: A37027
R;Yoshizuka, N.; Yoshimura, M.; Tsuchiya, S.; Okamoto, K.; Kobayashi, Y.; Osawa, T.
Cell. Immunol. 123, 212-225, 1989
A;Title: Macrophage chemotactic factor (MCF) produced by a human T cell hybridoma clone.
A;Reference number: A37027; MUID:89376581
A;Accession: A37027
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <YOS>


  Query Match             26.1%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   33.3%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLNVQD 6
            :|  :|
Db        1 WLGRZD 6


RESULT  19
B30572
T-cell receptor beta chain C region (CRTB29) - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Apr-1989 #sequence_revision 03-Apr-1989 #text_change 30-May-1997
C;Accession: B30572
R;Williams, C.B.; Gutman, G.A.
J. Immunol. 142, 1027-1035, 1989
A;Title: T cell receptor beta-chain genes in the rat. Availability and pattern of utiliz
A;Reference number: A30563; MUID:89110038
A;Accession: B30572
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-9 <WIL>
C;Keywords: T-cell receptor


  Query Match             26.1%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 QDL 7
            :||
Db        1 EDL 3


RESULT  20
HOROHA
proctolin - American cockroach
C;Species: Periplaneta americana (American cockroach)
C;Date: 29-Jul-1981 #sequence_revision 29-Jul-1981 #text_change 23-Aug-1996
C;Accession: A01644
R;Starratt, A.N.; Brown, B.E.
Life Sci. 17, 1253-1256, 1975
A;Title: Structure of the pentapeptide proctolin, a proposed neurotransmitter in insects
A;Reference number: A93048; MUID:76074708
A;Accession: A01644
A;Molecule type: protein
A;Residues: 1-5 <STA>
A;Note: the synthetic peptide had the same chromatographic, electrophoretic, and pharmac
R;O'Shea, M.; Adams, M.E.
Science 213, 567-569, 1981
A;Title: Pentapeptide (proctolin) associated with an identified neuron.
A;Reference number: A94260; MUID:81225865
A;Contents: annotation; biological source
C;Comment: This peptide is found in the lateral white neurons, which occur (in the cockr
innervate the striated hindgut muscles in insects and stimulate contraction of these mus
C;Superfamily: proctolin
```

```
C;Keywords: neuropeptide

  Query Match             23.9%;  Score 11;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YL 2
            ||
Db        2 YL 3


RESULT  21
B31836
20K protein - Rickettsia rickettsii (fragment)
C;Species: Rickettsia rickettsii
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 28-May-1999
C;Accession: B31836
R;Anderson, B.E.; Baumstark, B.R.; Bellini, W.J.
J. Bacteriol. 170, 4493-4500, 1988
A;Title: Expression of the gene encoding the 17-kilodalton antigen from Rickettsia ri
A;Reference number: A91885; MUID:89008059
A;Accession: B31836
A;Molecule type: DNA
A;Residues: 1-5 <AND>
A;Cross-references: GB:J03371; NID:g152455; PIDN:AAD15030.1; PID:g4262874


  Query Match             23.9%;  Score 11;  DB 2;  Length 5;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        6 DLN 8
            | |
Db        2 DTN 4


RESULT  22
A60411
proctolin - Atlantic horseshoe crab
C;Species: Limulus polyphemus (Atlantic horseshoe crab)
C;Date: 03-Feb-1993 #sequence_revision 03-Feb-1993 #text_change 17-Mar-1999
C;Accession: A60411
R;Groome, J.R.; Tillinghast, E.K.; Townley, M.A.; Vetrovs, A.; Watson III, W.H.; Hunt
Peptides 11, 205-211, 1990
A;Title: Identification of proctolin in the central nervous system of the horseshoe c
A;Reference number: A60411; MUID:90287800
A;Accession: A60411
A;Molecule type: protein
A;Residues: 1-5 <GRO>
C;Comment: This neuropeptide stimulates cardiac output and hindgut motility in the ho
C;Keywords: neuropeptide


  Query Match             23.9%;  Score 11;  DB 2;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YL 2
            ||
Db        2 YL 3


RESULT  23
PT0278
Ig heavy chain CRD3 region (clone 4-88) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C;Accession: PT0278
R;Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991
A;Title: Preferential utilization of specific immunoglobulin heavy chain diversity an
A;Reference number: PT0222; MUID:91108337
```

```
A;Accession: PT0278
A;Molecule type: DNA
A;Residues: 1-5 <YAM>
A;Experimental source: B lymphocyte
C;Keywords: heterotetramer; immunoglobulin


  Query Match             23.9%;  Score 11;  DB 2;  Length 5;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLNV 4
            |  |
Db        1 YFGV 4


RESULT  24
B44510
hypothetical protein C (mleR 3' region) - Lactococcus lactis (fragment)
C;Species: Lactococcus lactis
C;Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 30-Sep-1993
C;Accession: B44510
R;Renault, P.; Gaillardin, C.; Heslot, H.
J. Bacteriol. 171, 3108-3114, 1989
A;Title: Product of the Lactococcus lactis gene required for malolactic fermentation is
A;Reference number: A44510; MUID:89255069
A;Accession: B44510
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-6 <REN>
A;Cross-references: EMBL:M90762


  Query Match             23.9%;  Score 11;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YL 2
            ||
Db        5 YL 6


RESULT  25
JH0784
neuropeptide TE-6 - pig roundworm (fragment)
C;Species: Ascaris suum (pig roundworm)
C;Date: 10-Jun-1993 #sequence_revision 10-Jun-1993 #text_change 28-Apr-1995
C;Accession: JH0784
R;Smart, D.; Shaw, C.; Curry, W.J.; Johnston, C.F.; Thim, L.; Halton, D.W.; Buchanan, K.
Biochem. Biophys. Res. Commun. 187, 1323-1329, 1992
A;Title: The primary structure of TE-6: a novel neuropeptide from the nematode Ascaris s
A;Reference number: JH0784; MUID:93038603
A;Accession: JH0784
A;Molecule type: protein
A;Residues: 1-6 <SMA>
A;Experimental source: gonoduct
C;Keywords: neuropeptide


  Query Match             23.9%;  Score 11;  DB 2;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 QDL 7
            |:|
Db        3 QEL 5


RESULT  26
S78024
ribosomal protein YmL21, mitochondrial - yeast (Saccharomyces cerevisiae) (fragment)
C;Species: Saccharomyces cerevisiae
C;Date: 09-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 14-Nov-1997
C;Accession: S78024
R;Kitakawa, M.; Graack, H.R.; Grohmann, L.; Goldschmidt-Reisin, S.; Herfurth, E.; Wit
Eur. J. Biochem. 245, 449-456, 1997
A;Title: Identification and characterization of the genes for mitochondrial ribosomal
A;Reference number: S78018; MUID:97296414
A;Accession: S78024
A;Molecule type: protein
A;Residues: 1-7 <KIT>
C;Genetics:
A;Genome: nuclear
C;Keywords: mitochondrion; protein biosynthesis; ribosome


  Query Match             23.9%;  Score 11;  DB 2;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YL 2
            ||
Db        1 YL 2


RESULT  27
A11483
aspartate transaminase (EC 2.6.1.1), mitochondrial - sheep (fragment)
N;Alternate names: aspartate aminotransferase, mitochondrial
C;Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C;Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 20-Aug-1999
C;Accession: A11483
R;Campos-Cavieres, M.; Milstein, C.P.
Biochem. J. 147, 275-281, 1975
A;Title: The sequences of the coenzyme-binding peptide in the cytoplasmic and the mit
A;Reference number: A11483; MUID:76039441
A;Accession: A11483
A;Molecule type: protein
A;Residues: 1-7 <CAM>
A;Experimental source: liver
C;Keywords: aminotransferase; mitochondrion; phosphoprotein; pyridoxal phosphate
F;2/Binding site: pyridoxal phosphate (Lys) (covalent) #status experimental


  Query Match             23.9%;  Score 11;  DB 2;  Length 7;
  Best Local Similarity   20.0%;  Pred. No. 1.8e+05;
  Matches    1;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        5 QDLNL 9
            ::: |
Db        2 KBMGL 6


RESULT  28
PQ0726
unidentified 4.5/45K [imported] - rice (fragment)
C;Species: Oryza sativa (rice)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: PQ0726
R;Komatsu, S.; Kajiwara, H.; Hirano, H.
Theor. Appl. Genet. 86, 935-942, 1993
A;Title: A rice protein library; a data-file of rice proteins separated by two-dimens
A;Reference number: PQ0696
A;Accession: PQ0726
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-8 <KOM>


  Query Match             23.9%;  Score 11;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 QD 6
            ||
Db        4 QD 5
```

```
RESULT  29
PT0311
Ig heavy chain CRD3 region (clone 6-100) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C;Accession: PT0311
R;Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991
A;Title: Preferential utilization of specific immunoglobulin heavy chain diversity and j
A;Reference number: PT0222; MUID:91108337
A;Accession: PT0311
A;Molecule type: DNA
A;Residues: 1-8 <YAM>
A;Experimental source: B lymphocyte
C;Keywords: heterotetramer; immunoglobulin


  Query Match             23.9%;  Score 11;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YL 2
          ||
Db      7 YL 8


RESULT  30
A42689
major postsynaptic density protein - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 26-May-1994
C;Accession: A42689
R;Wu, K.; Huang, Y.; Adler, J.; Black, I.B.
Proc. Natl. Acad. Sci. U.S.A. 89, 3015-3019, 1992
A;Title: On the identity of the major postsynaptic density protein.
A;Reference number: A42689; MUID:92212958
A;Accession: A42689
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-8 <WUA>


  Query Match             23.9%;  Score 11;  DB 2;  Length 8;
  Best Local Similarity   28.6%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 LNVQDLN 8
          | | :::
Db      2 LKVPNIS 8


RESULT  31
A44873
caldesmon - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 31-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 31-Oct-1997
C;Accession: A44873
R;Ikebe, M.; Hornick, T.
Arch. Biochem. Biophys. 288, 538-542, 1991
A;Title: Determination of the phosphorylation sites of smooth muscle caldesmon by protei
A;Reference number: A44873; MUID:91378498
A;Accession: A44873
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <IKE>
A;Experimental source: skeletal myosin
A;Note: sequence extracted from NCBI backbone (NCBIP:63199)
C;Superfamily: caldesmon


  Query Match             23.9%;  Score 11;  DB 2;  Length 9;
```

```
  Best Local Similarity   20.0%;  Pred. No. 1.8e+05;
  Matches    1;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 LNVQD 6
          | :::
Db      4 LKIEE 8


RESULT  32
A61230
calsequestrin, cardiac and slow skeletal muscle - northern leopard frog (fragment)
N;Alternate names: 58K dihydropyridine-binding protein; aspartactin; calmitine; lamin
C;Species: Rana pipiens (northern leopard frog)
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 16-Jul-1999
C;Accession: A61230
R;McLeod, A.G.; Shen, A.C.Y.; Campbell, K.P.; Michalak, M.; Jorgensen, A.O.
Circ. Res. 69, 344-359, 1991
A;Title: Frog cardiac calsequestrin. Identification, characterization, and subcellula
rdium.
A;Reference number: A61230; MUID:91316784
A;Accession: A61230
A;Molecule type: protein
A;Residues: 1-9 <MCL>
C;Comment: Calsequestrin is a high-capacity and moderate-affinity calcium binding pro
C;Comment: Calsequestrin acts as a calcium buffer, and the release of calcium bound t
C;Comment: The fast skeletal muscle isoform of calsequestrin can be phosphorylated in
C;Superfamily: calsequestrin
C;Keywords: calcium binding; cardiac muscle; glycoprotein; heart; phosphoprotein; ske


  Query Match             23.9%;  Score 11;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      5 QDLN 8
           : ||
Db      2 EGLN 5


RESULT  33
A60108
exotoxin A - Streptococcus pyogenes (strain C203.S) (fragment)
N;Alternate names: blastogen A; scarlet fever toxin
C;Species: Streptococcus pyogenes
C;Date: 10-Nov-1992 #sequence_revision 10-Nov-1992 #text_change 07-Feb-1997
C;Accession: A60108
R;Schlievert, P.M.; Gray, E.D.
Infect. Immun. 57, 1865-1867, 1989
A;Title: Group A streptococcal pyrogenic exotoxin (scarlet fever toxin) type A and bl
A;Reference number: A60108; MUID:89254013
A;Accession: A60108
A;Molecule type: protein
A;Residues: 1-9 <SCH>
C;Keywords: exotoxin


  Query Match             23.9%;  Score 11;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 QD 6
          ||
Db      1 QD 2


RESULT  34
A41978
calliFMRFamide 1 - bluebottle fly (Calliphora vomitoria)
C;Species: Calliphora vomitoria
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C;Accession: A41978
R;Duve, H.; Johnsen, A.H.; Sewell, J.C.; Scott, A.G.; Orchard, I.; Rehfeld, J.F.; Tho
Proc. Natl. Acad. Sci. U.S.A. 89, 2326-2330, 1992
```

```
A;Title: Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH-2 neuropeptides (desi
A;Reference number: A41978; MUID:92196111
A;Accession: A41978
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <DUV>
C;Keywords: amidated carboxyl end; neuropeptide
F;9/Modified site: amidated carboxyl end (Phe) #status experimental


  Query Match             23.9%;  Score 11;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 QD 6
            ||
Db        4 QD 5


RESULT  35
D44787
calliFMRFamide 13 - bluebottle fly (Calliphora vomitoria)
C;Species: Calliphora vomitoria
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C;Accession: D44787
R;Duve, H.; Johnsen, A.H.; Sewell, J.C.; Scott, A.G.; Orchard, I.; Rehfeld, J.F.; Thorpe
Proc. Natl. Acad. Sci. U.S.A. 89, 2326-2330, 1992
A;Title: Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH-2 neuropeptides (desi
A;Reference number: A41978; MUID:92196111
A;Accession: D44787
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <DUV>
C;Keywords: amidated carboxyl end; neuropeptide
F;9/Modified site: amidated carboxyl end (Phe) #status experimental


  Query Match             23.9%;  Score 11;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 QD 6
            ||
Db        3 QD 4


Search completed: December 16, 2000, 03:35:13
Job time: 5646 sec
```

```
                    GenCore version 4.5
            Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 04:23:31 ; Search time 68.03 Seconds
                                                (without alignments)
                                                4.227 Million cell updates/sec

Title:           US-09-529-121-9
Perfect score:   46
Sequence:        1 YLNVQDLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters:      213

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       SwissProt_39:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                %
Result          Query
  No.   Score   Match Length DB  ID                         Description
-------------------------------------------------------------------------------
    1     21    45.7       9  1  LMIP_LOCMI                 P31799 locusta mig
    2     15    32.6       9  1  DCML_PSECF                 P19913 pseudomonas
    3     15    32.6       9  1  OXYA_SCYCA                 P42996 scyliorhinu
    4     15    32.6       9  1  OXYA_SQUAC                 P42999 squalus aca
    5     15    32.6       9  1  ULAH_HUMAN                 P31934 homo sapien
    6     14    30.4       8  1  PLP_BRANA                  P81707 brassica na
    7     14    30.4       9  1  UF02_MOUSE                 P38640 mus musculu
    8     13    28.3       9  1  OXYF_SCYCA                 P42997 scyliorhinu
    9     13    28.3       9  1  UPA6_HUMAN                 P30092 homo sapien
   10     12    26.1       8  1  CCKN_MACEU                 P30369 macropus eu
   11     12    26.1       8  1  UPA1_HUMAN                 P30087 homo sapien
   12     11    23.9       4  1  FAR3_HIRME                 P42562 hirudo medi
   13     11    23.9       5  1  PRCT_PERAM                 P01373 periplaneta
   14     11    23.9       7  1  FAR2_ASCSU                 P31890 ascaris suu
   15     11    23.9       7  1  GFRP_MOUSE                 P99025 mus musculu
   16     11    23.9       9  1  D1_NEPNO                   P24816 nephrops no
   17     11    23.9       9  1  FAR1_CALVO                 P41856 calliphora
   18     11    23.9       9  1  FAR2_CALVO                 P41857 calliphora
   19     11    23.9       9  1  FAR3_CALVO                 P41858 calliphora
   20     11    23.9       9  1  FAR4_CALVO                 P41859 calliphora
   21     11    23.9       9  1  FAR5_CALVO                 P41860 calliphora
   22     11    23.9       9  1  FAR6_CALVO                 P41861 calliphora
   23     11    23.9       9  1  FAR7_CALVO                 P41862 calliphora
   24     11    23.9       9  1  FARD_CALVO                 P41868 calliphora
   25     10    21.7       3  1  LUXE_VIBFI                 P24272 vibrio fisc
   26     10    21.7       6  1  TMOF_SARBU                 P41495 sarcophaga
   27     10    21.7       6  1  TRPI_PSEPU                 P36414 pseudomonas
   28     10    21.7       7  1  ALL7_CYDPO                 P82158 cydia pomon
   29     10    21.7       7  1  FAR2_PROCL                 P38498 procambarus
   30     10    21.7       7  1  UN06_PINPS                 P81675 pinus pinas
   31     10    21.7       8  1  ACT_CARMA                  P80709 carcinus ma
   32     10    21.7       8  1  AKH_MELML                  P25423 melolontha
   33     10    21.7       8  1  ALL5_CYDPO                 P82156 cydia pomon
```

```
   34     10    21.7       8  1  ANG2_BOTJA                 Q10582 bothrops ja
   35     10    21.7       8  1  B44K_PORGI                 P81886 porphyromon
   36     10    21.7       8  1  COXG_RAT                   P80430 rattus norv
   37     10    21.7       8  1  FAR3_HOMAM                 P41486 homarus ame
   38     10    21.7       8  1  FUSS_FUSSO                 P81010 fusarium so
   39     10    21.7       8  1  HTF_TENMO                  P25419 tenebrio mo
   40     10    21.7       8  1  LCK3_LEUMA                 P21142 leucophaea
   41     10    21.7       8  1  RPCH_PANBO                 P08939 pandalus bo
   42     10    21.7       9  1  FIBB_MACFU                 P19345 macaca fusc
   43     10    21.7       9  1  ISOT_CYPCA                 P42993 cyprinus ca
   44     10    21.7       9  1  OXYT_RABIT                 P32878 oryctolagus
   45     10    21.7       9  1  OXYT_RAJCL                 P42994 raja clavat
   46      9    19.6       4  1  FAR4_HIRME                 P42563 hirudo medi
   47      9    19.6       5  1  FARP_ARTTR                 P41853 artioposthi
   48      9    19.6       6  1  UN06_CLOPA                 P81351 clostridium
   49      9    19.6       7  1  FAR3_HAECO                 P81298 haemonchus
   50      9    19.6       7  1  FAR3_PANRE                 P41874 panagrellus
   51      9    19.6       7  1  FAR5_HIRME                 P42564 hirudo medi
   52      9    19.6       8  1  CLP_THICU                  P80488 thiobacillu
   53      9    19.6       8  1  FAR1_PANRE                 P41872 panagrellus
   54      9    19.6       9  1  FAR2_PANRE                 P41873 panagrellus
   55      9    19.6       9  1  FIBB_PAPAN                 P19344 papio anubi
   56      9    19.6       9  1  FIBB_PAPHA                 P19343 papio hamad
   57      9    19.6       9  1  FIBB_THEGE                 P19342 theropithec
   58      9    19.6       9  1  FLA2_TREHY                 P80159 treponema h
   59      9    19.6       9  1  OXYT_BUFRE                 P42995 bufo regula
   60      9    19.6       9  1  OXYT_CYPCA                 P23879 cyprinus ca
   61      9    19.6       9  1  OXYV_SQUAC                 P43000 squalus aca
   62      9    19.6       9  1  ULAE_HUMAN                 P31931 homo sapien
   63      8    17.4       5  1  BIOA_CITFR                 P13071 citrobacter
   64      8    17.4       5  1  BIOA_SALTY                 P12677 salmonella
   65      8    17.4       7  1  CARP_MYTED                 P10420 mytilus edu
   66      8    17.4       7  1  IGAO_DACDE                 P06294 dactylium d
   67      8    17.4       7  1  LANC_CARUI                 P36960 carnobacter
   68      8    17.4       7  1  MYOM_APLCA                 P15513 aplysia cal
   69      8    17.4       8  1  AL12_CARMA                 P81815 carcinus ma
   70      8    17.4       8  1  AL18_CARMA                 P81821 carcinus ma
   71      8    17.4       8  1  ALL5_CALVO                 P41841 calliphora
   72      8    17.4       8  1  CPD1_ENTFA                 P13269 enterococcu
   73      8    17.4       9  1  BUK_CLOPA                  P81337 clostridium
   74      8    17.4       9  1  CCAP_CARMA                 P38556 carcinus ma
   75      8    17.4       9  1  LPCA_STAAU                 P36884 staphylococ


                                   ALIGNMENTS


RESULT   1
LMIP_LOCMI
ID   LMIP_LOCMI     STANDARD;      PRT;     9 AA.
AC   P31799;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-OCT-1993 (Rel. 27, Last annotation update)
DE   LOCUSTAMYOINHIBITING PEPTIDE (LOM-MIP).
OS   Locusta migratoria (Migratory locust).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC   Acridomorpha; Acridoidea; Acrididae; Locusta.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92179466.
RA   Schoofs L., Holman G.M., Hayes T.K., Nachman R.J., de Loof A.;
RT   "Isolation, identification and synthesis of locustamyoinhibiting
RT   peptide (LOM-MIP), a novel biologically active neuropeptide from
RT   Locusta migratoria.";
RL   Regul. Pept. 36:111-119(1991).
CC   -!- FUNCTION: SUPPRESSES SPONTANEOUS CONTRACTIONS OF THE HINDGUT AND
CC       OVIDUCT.
CC   -!- TISSUE SPECIFICITY: NEURONS LOCATED IN TWO VENTRAL CELL CLUSTERS
CC       IN THE SUBOESOPHAGEAL GANGLION.
DR   PIR; A60065; AKLQIM.
KW   Amidation; Neuropeptide.
```

```
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1060 MW;  387D7DD4472AB6C3 CRC64;


  Query Match             45.7%;  Score 21;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 QDLN 8
            ||||
Db        3 QDLN 6


RESULT   2
DCML_PSECF
ID   DCML_PSECF     STANDARD;      PRT;     9 AA.
AC   P19913;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-DEC-1992 (Rel. 24, Last annotation update)
DE   CARBON MONOXIDE DEHYDROGENASE LARGE CHAIN (EC 1.2.99.2) (FRAGMENT).
OS   Pseudomonas carboxydoflava.
OC   Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
OC   Hydrogenophaga.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 90055678.
RA   Kraut M., Hugendieck I., Herwig S., Meyer O.;
RT   "Homology and distribution of CO dehydrogenase structural genes in
RT   carboxydotrophic bacteria.";
RL   Arch. Microbiol. 152:335-341(1989).
CC   -!- CATALYTIC ACTIVITY: CO + H(2)O + ACCEPTOR = CO(2) + REDUCED
CC       ACCEPTOR.
CC   -!- COFACTOR: MOLYBDENUM.
CC   -!- SUBUNIT: CONSISTS OF THREE POLYPEPTIDE CHAINS: LARGE, MEDIUM, AND
CC       SMALL.
DR   PIR; PL0139; PL0139.
KW   Oxidoreductase; Molybdenum.
FT   VARIANT       2      2       N -> M.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  974 MW;  0224DAB6C2D76DD4 CRC64;


  Query Match             32.6%;  Score 15;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        4 VQD 6
            |||
Db        5 VQD 7


RESULT   3
OXYA_SCYCA
ID   OXYA_SCYCA     STANDARD;      PRT;     9 AA.
AC   P42996;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ASVATOCIN.
OS   Scyliorhinus canicula (Spotted dogfish) (Spotted catshark).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Galeomorphii; Galeoidea; Carcharhiniformes;
OC   Scyliorhinidae; Scyliorhinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PITUITARY;
RX   MEDLINE; 95062247.
RA   Chauvet J., Rouille Y., Chauveau C., Chauvet M.-T., Acher R.;
RT   "Special evolution of neurohypophysial hormones in cartilaginous
RT   fishes: asvatocin and phasvatocin, two oxytocin-like peptides
RT   isolated from the spotted dogfish (Scyliorhinus caniculus).";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:11266-11270(1994).
CC   -!- FUNCTION: DISPLAYS OXYTOCIC ACTIVITY ON RAT UTERUS.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  982 MW;  17EDD76EB444404B CRC64;


  Query Match             32.6%;  Score 15;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLN 3
            |:|
Db        2 YIN 4


RESULT   4
OXYA_SQUAC
ID   OXYA_SQUAC     STANDARD;      PRT;     9 AA.
AC   P42999;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ASPARTOCIN (ASPARGTOCIN).
OS   Squalus acanthias (Spiny dogfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Squaloidei; Squalidae; Squalus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 73031727.
RA   Acher R., Chauvet J., Chauvet M.-T.;
RT   "Phylogeny of the neurohypophysial hormones. Two new active peptides
RT   isolated from a cartilaginous fish, Squalus acanthias.";
RL   Eur. J. Biochem. 29:12-19(1972).
RN   [2]
RP   SEQUENCE.
RX   MEDLINE; 72128038.
RA   Acher R., Chauvet J., Chauvet M.-T., Fontaine M.;
RT   "Identification of 2 new neurohypophyseal hormones, valitocin (Val8-
RT   oxytocin) and aspartocin (Asn4-oxytocin) in a selachian fish, the
RT   spiny dog-fish (Squalus acanthias).";
RL   C. R. Acad. Sci., D, Sci. Nat. 274:313-316(1972).
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  996 MW;  17F8376EB444404B CRC64;


  Query Match             32.6%;  Score 15;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLN 3
            |:|
Db        2 YIN 4


RESULT   5
ULAH_HUMAN
ID   ULAH_HUMAN     STANDARD;      PRT;     9 AA.
AC   P31934;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE OF LIVER TISSUE (SPOT 94) (FRAGMENT).
```

```
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LIVER;
RX   MEDLINE; 94147969.
RA   Hughes G.J., Frutiger S., Paquet N., Pasquali C., Sanchez J.-C.,
RA   Tissot J.-D., Bairoch A., Appel R.D., Hochstrasser D.F.;
RT   "Human liver protein map: update 1993.";
RL   Electrophoresis 14:1216-1222(1993).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 5.5, ITS MW IS: 12 KDA.
DR   SWISS-2DPAGE; P31934; HUMAN.
FT   NON_TER       1      1
FT   VARIANT       3      3       A -> L.
FT                                /FTId=VAR_000001.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  990 MW;  F1CC7B172AB6DDCA CRC64;


  Query Match             32.6%;  Score 15;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 QDL 7
           |||
Db       4 QDL 6


RESULT   6
PLP_BRANA
ID   PLP_BRANA      STANDARD;      PRT;     8 AA.
AC   P81707;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PLASTIDIAL LIPID-ASSOCIATED PROTEIN (FRAGMENT).
OS   Brassica napus (Rape).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Brassica.
RN   [1]
RP   SEQUENCE.
RC   STRAIN=CV. TOPAZ; TISSUE=TAPETUM;
RX   MEDLINE; 99349136.
RA   Hernandez-Pinzon I., Ross J.H.E., Barnes K.A., Damant A.P.,
RA   Murphy D.J.;
RT   "Composition and role of tapetal lipid bodies in the biogenesis of the
RT   pollen coat of Brassica napus.";
RL   Planta 208:588-598(1999).
CC   -!- FUNCTION: MAY PLAY A STRUCTURAL ROLE IN THE ELAIOPLAST, A TAPETUM-
CC       SPECIFIC PLASTIDIAL LIPID ORGANELLE.
CC   -!- TISSUE SPECIFICITY: TAPETUM OF ANTHERS.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  989 MW;  9D7B1AA452CAA042 CRC64;


  Query Match             30.4%;  Score 14;  DB 1;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 8.8e+04;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 VQDLN 8
           | |:|
Db       1 VIDVN 5


RESULT   7
UF02_MOUSE
ID   UF02_MOUSE     STANDARD;      PRT;     9 AA.
AC   P38640;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE OF FIBROBLASTS (P32) (FRAGMENT).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=FIBROBLAST;
RX   MEDLINE; 95009907.
RA   Merrick B.A., Patterson R.M., Wichter L.L., He C., Selkirk J.K.;
RT   "Separation and sequencing of familiar and novel murine proteins
RT   using preparative two-dimensional gel electrophoresis.";
RL   Electrophoresis 15:735-745(1994).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 6.0, ITS MW IS: 32 KDA.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1102 MW;  7E73EAB6D05B1AAB CRC64;

  Query Match             30.4%;  Score 14;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 VQD 6
           :||
Db       5 IQD 7


RESULT   8
OXYF_SCYCA
ID   OXYF_SCYCA     STANDARD;      PRT;     9 AA.
AC   P42997;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   PHASVATOCIN.
OS   Scyliorhinus canicula (Spotted dogfish) (Spotted catshark).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Galeomorphii; Galeoidea; Carcharhiniformes;
OC   Scyliorhinidae; Scyliorhinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PITUITARY;
RX   MEDLINE; 95062247.
RA   Chauvet J., Rouille Y., Chauveau C., Chauvet M.-T., Acher R.;
RT   "Special evolution of neurohypophysial hormones in cartilaginous
RT   fishes: asvatocin and phasvatocin, two oxytocin-like peptides
RT   isolated from the spotted dogfish (Scyliorhinus caniculus).";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:11266-11270(1994).
CC   -!- FUNCTION: DISPLAYS OXYTOCIC ACTIVITY ON RAT UTERUS.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1016 MW;  17EDD76EB44449DB CRC64;


  Query Match             28.3%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLN 3
           | |
Db       2 YFN 4


RESULT   9
UPA6_HUMAN
ID   UPA6_HUMAN     STANDARD;      PRT;     9 AA.
AC   P30092;
```

```
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE OF PLASMA (SPOT 14) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PLASMA;
RX   MEDLINE; 93092937.
RA   Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,
RA   Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,
RA   Hochstrasser D.F.;
RT   "Plasma protein map: an update by microsequencing.";
RL   Electrophoresis 13:707-714(1992).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 5, ITS MW IS: 48 KDA.
DR   SWISS-2DPAGE; P30092; HUMAN.
FT   NON_TER       1      1
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  935 MW;  5282F2CAA8676447 CRC64;

  Query Match             28.3%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        2 LNVQDL 7
            || |:
Db        2 LNPGDV 7


RESULT  10
CCKN_MACEU
ID   CCKN_MACEU     STANDARD;      PRT;     8 AA.
AC   P30369;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   CHOLECYSTOKININ (CCK).
GN   CCK.
OS   Macropus eugenii (Tammar wallaby), and
OS   Dasyurus viverrinus (Southeastern quoll).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=BRAIN;
RX   MEDLINE; 88234141.
RA   Fan Z.W., Eng J., Shaw G., Yalow R.S.;
RT   "Cholecystokinin octapeptide purified from brains of Australian
RT   marsupials.";
RL   Peptides 9:429-431(1988).
CC   -!- FUNCTION: THIS PEPTIDE HORMONE INDUCES GALL BLADDER CONTRACTION
CC       AND THE RELEASE OF PANCREATIC ENZYMES IN THE GUT. ITS FUNCTION
CC       IN THE BRAIN IS NOT CLEAR.
CC   -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
DR   PIR; A43001; A43001.
DR   PIR; PQ0012; PQ0012.
DR   INTERPRO; IPR001651; -.
DR   PROSITE; PS00259; GASTRIN; 1.
KW   Amidation; Sulfatation; Hormone.
FT   MOD_RES       2      2       SULFATATION.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  1064 MW;  DDCAA68378768B5A CRC64;

  Query Match             26.1%;  Score 12;  DB 1;  Length 8;
  Best Local Similarity   33.3%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        1 YLNVQD 6
```

```
              |:  |
Db        2 YMGWMD 7


RESULT  11
UPA1_HUMAN
ID   UPA1_HUMAN     STANDARD;      PRT;     8 AA.
AC   P30087;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE OF PLASMA (SPOT 2) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PLASMA;
RX   MEDLINE; 93092937.
RA   Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,
RA   Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,
RA   Hochstrasser D.F.;
RT   "Plasma protein map: an update by microsequencing.";
RL   Electrophoresis 13:707-714(1992).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 4.9, ITS MW IS: 65 KDA.
DR   SWISS-2DPAGE; P30087; HUMAN.
FT   NON_TER       1      1
FT   UNSURE        8      8
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  944 MW;  C01772C455BB06DA CRC64;

  Query Match             26.1%;  Score 12;  DB 1;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        5 QDLNL 9
            |: |:
Db        2 QESNV 6


RESULT  12
FAR3_HIRME
ID   FAR3_HIRME     STANDARD;      PRT;     4 AA.
AC   P42562;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE YLRF-AMIDE.
OS   Hirudo medicinalis (Medicinal leech).
OC   Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;
OC   Arynchobdellida; Hirudiniformes; Hirudinidae; Hirudo.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92195954.
RA   Evans B.D., Pohl J., Kartsonis M.A., Calabrese R.L.;
RT   "Identification of RFamide neuropeptides in the medicinal leech.";
RL   Peptides 12:897-908(1991).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       4      4       AMIDATION.
SQ   SEQUENCE   4 AA;  598 MW;  69D4073B30000000 CRC64;

  Query Match             23.9%;  Score 11;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YL 2
            ||
Db        1 YL 2
```

```
RESULT  13
PRCT_PERAM
ID   PRCT_PERAM     STANDARD;      PRT;     5 AA.
AC   P01373;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   PROCTOLIN.
OS   Periplaneta americana (American cockroach),
OS   Limulus polyphemus (Atlantic horseshoe crab), and
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blattoidea; Blattidae; Periplaneta.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=P.AMERICANA;
RX   MEDLINE; 76074708.
RA   Starratt A.N., Brown B.E.;
RT   "Structure of the pentapeptide proctolin, a proposed neurotransmitter
RT   in insects.";
RL   Life Sci. 17:1253-1256(1975).
RN   [2]
RP   BIOLOGICAL SOURCE.
RC   SPECIES=P.AMERICANA;
RX   MEDLINE; 81225865.
RA   O'Shea M., Adams M.E.;
RT   "Pentapeptide (proctolin) associated with an identified neuron.";
RL   Science 213:567-569(1981).
RN   [3]
RP   SEQUENCE.
RC   SPECIES=L.POLYPHEMUS;
RX   MEDLINE; 90287800.
RA   Groome J.R., Tillinghast E.K., Townley M.A., Vetrovs A.,
RA   Watson W.H. III, Hunt D.F., Griffin P.R., Alexander J.E.,
RA   Shabanowitz J.;
RT   "Identification of proctolin in the central nervous system of the
RT   horseshoe crab, Limulus polyphemus.";
RL   Peptides 11:205-211(1990).
RN   [4]
RP   SEQUENCE.
RC   SPECIES=C.MAENAS;
RX   MEDLINE; 86232789.
RA   Stangler J., Dircksen H., Keller R.;
RT   "Identification and immunocytochemical localization of proctolin in
RT   pericardial organs of the shore crab, Carcinus maenas.";
RL   Peptides 7:67-72(1986).
CC   -!- FUNCTION: STIMULATES CARDIAC OUTPUT AND HINDGUT MOTILITY,
CC       MODULATES VISCERAL AND SKELETAL MUSCLE IN MANY ARTHROPODS.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE LATERAL WHITE NEURONS AND IN
CC       THE CRAB PERICARDIAL ORGANS.
DR   PIR; A01644; HOROHA.
DR   PIR; A60411; A60411.
KW   Neuropeptide.
SQ   SEQUENCE   5 AA;  649 MW;  71B7673B44600000 CRC64;

  Query Match             23.9%;  Score 11;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YL 2
           ||
Db       2 YL 3

RESULT  14
FAR2_ASCSU
ID   FAR2_ASCSU     STANDARD;      PRT;     7 AA.
AC   P31890;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE AF2.
OS   Ascaris suum (Pig roundworm) (Ascaris lumbricoides), and
OS   Panagrellus redivivus.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC   Ascarididae; Ascaris.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=A.SUUM;
RX   MEDLINE; 93324431.
RA   Cowden C., Stretton A.O.W.;
RT   "AF2, an Ascaris neuropeptide: isolation, sequence, and bioactivity.";
RL   Peptides 14:423-430(1993).
RN   [2]
RP   SEQUENCE.
RC   SPECIES=P.REDIVIVUS;
RX   MEDLINE; 95060998.
RA   Maule A.G., Shaw C., Bowman J.W.;
RT   "The FMRFamide-like neuropeptide AF2 (Ascaris suum) is present in the
RT   free-living nematode, Panagrellus redivivus (Nematoda, Rhabditida).";
RL   Parasitology 109:351-356(1994).
CC   -!- FUNCTION: HAS EFFECTS ON MUSCLE TENSION.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE NERVE CORDS AND A VARIETY OF
CC       GANGLIA PARTICULARLY IN THE ANTERIOR REGIONS.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES      7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  992 MW;  69D4073B5B11E350 CRC64;

  Query Match             23.9%;  Score 11;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YL 2
           ||
Db       4 YL 5

RESULT  15
GFRP_MOUSE
ID   GFRP_MOUSE     STANDARD;      PRT;     7 AA.
AC   P99025;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GTP CYCLOHYDROLASE I FEEDBACK REGULATORY PROTEIN (P35) (FRAGMENT).
GN   GFRP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LIVER;
RA   Sanchez J.-C., Rouge V., Frutiger S., Hughes G., Yan J.X.,
RA   Hoogland C., Appel R.D., Binz P.-A., Hochstrasser D.F.,
RA   Cowthorne M.;
RL   Submitted (AUG-1998) to the SWISS-PROT data bank.
CC   -!- FUNCTION: MEDIATES TETRAHYDROBIOPTERIN INHIBITION OF GTP
CC       CYCLOHYDROLASE I. THIS INHIBITION IS REVERSED BY L-PHENYLALANINE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
DR   SWISS-2DPAGE; P99025; MOUSE.
FT   INIT_MET     0      0
FT   NON_TER      7      7
SQ   SEQUENCE   7 AA;  806 MW;  71B5B057273B4700 CRC64;

  Query Match             23.9%;  Score 11;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 YL 2
           ||
Db       2 YL 3


RESULT  16
D1_NEPNO
ID   D1_NEPNO       STANDARD;      PRT;     9 AA.
AC   P24816;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   01-MAR-1992 (Rel. 21, Last annotation update)
DE   GASTRIN/CHOLECYSTOKININ-LIKE PEPTIDE D1.
OS   Nephrops norvegicus (Norway lobster).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC   Nephropoidea; Nephropidae; Nephrops.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=STOMACH;
RX   MEDLINE; 92082847.
RA   Favrel P., Kegel G., Sedlmeier D., Keller R., van Wormhoudt A.;
RT   "Structure and biological activity of crustacean gastrointestinal
RT   peptides identified with antibodies to gastrin/cholecystokinin.";
RL   Biochimie 73:1233-1239(1991).
CC   -!- FUNCTION: MAY CONTROL DIGESTION PROCESSES IN CRUSTACEANS.
CC   -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
DR   PIR; A48398; A48398.
KW   Hormone.
SQ   SEQUENCE   9 AA;  1038 MW;  60EC79CAB6D8787B CRC64;


  Query Match             23.9%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 QD 6
           ||
Db       5 QD 6


RESULT  17
FAR1_CALVO
ID   FAR1_CALVO     STANDARD;      PRT;     9 AA.
AC   P41856;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 1.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- FUNCTION: ABLE TO INDUCE FLUID SECRETION FROM THE ISOLATED
CC       SALIVARY GLAND OF CALLIPHORA.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; A41978; A41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1169 MW;  29D00699CAB6C6C7 CRC64;

  Query Match             23.9%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 QD 6
           ||
Db       4 QD 5


RESULT  18
FAR2_CALVO
ID   FAR2_CALVO     STANDARD;      PRT;     9 AA.
AC   P41857;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 2.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- FUNCTION: ABLE TO INDUCE FLUID SECRETION FROM THE ISOLATED
CC       SALIVARY GLAND OF CALLIPHORA.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; B41978; B41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1128 MW;  29D00699CAB6C5A7 CRC64;

  Query Match             23.9%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 QD 6
           ||
Db       4 QD 5


RESULT  19
FAR3_CALVO
ID   FAR3_CALVO     STANDARD;      PRT;     9 AA.
AC   P41858;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 3.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
```

```
CC   -!- FUNCTION: ABLE TO INDUCE FLUID SECRETION FROM THE ISOLATED
CC       SALIVARY GLAND OF CALLIPHORA.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; C41978; C41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1114 MW;  2F0B0699CAB6C5A7 CRC64;

  Query Match            23.9%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 QD 6
           ||
Db       4 QD 5


RESULT  20
FAR4_CALVO
ID   FAR4_CALVO     STANDARD;       PRT;     9 AA.
AC   P41859;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 4.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; D41978; D41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1182 MW;  31730699CAB6D457 CRC64;

  Query Match            23.9%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 QD 6
           ||
Db       4 QD 5


RESULT  21
FAR5_CALVO
ID   FAR5_CALVO     STANDARD;       PRT;     9 AA.
AC   P41860;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 5.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
```

```
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; E41978; E41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1068 MW;  39D10699CAB6D867 CRC64;

  Query Match            23.9%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 QD 6
           ||
Db       4 QD 5


RESULT  22
FAR6_CALVO
ID   FAR6_CALVO     STANDARD;       PRT;     9 AA.
AC   P41861;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 6.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; F41978; F41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1058 MW;  96D10699CAB6D865 CRC64;

  Query Match            23.9%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 QD 6
           ||
Db       4 QD 5


RESULT  23
FAR7_CALVO
ID   FAR7_CALVO     STANDARD;       PRT;     9 AA.
AC   P41862;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 7.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
```

```
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; G41978; G41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1081 MW;  E1D10699CAB6D86A CRC64;

  Query Match             23.9%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 QD 6
            ||
Db        4 QD 5


RESULT  24
FARD_CALVO
ID   FARD_CALVO     STANDARD;      PRT;     9 AA.
AC   P41868;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 13.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; D44787; D44787.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1028 MW;  22D10699C87AB6D8 CRC64;

  Query Match             23.9%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 QD 6
            ||
Db        3 QD 4


RESULT  25
LUXE_VIBFI
ID   LUXE_VIBFI     STANDARD;      PRT;     3 AA.
AC   P24272;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   LONG-CHAIN-FATTY-ACID--LUCIFERIN-COMPONENT LIGASE (EC 6.2.1.19)
DE   (ACYL-PROTEIN SYNTHETASE) (FRAGMENT).
GN   LUXE.
OS   Vibrio fischeri.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91072226.
RA   Swartzman E., Kapoor S., Graham A.F., Meighen E.A.;
RT   "A new Vibrio fischeri lux gene precedes a bidirectional termination
RT   site for the lux operon.";
RL   J. Bacteriol. 172:6797-6802(1990).
CC   -!- FUNCTION: ACYL-PROTEIN SYNTHETASE ACTIVATES TETRADECANOIC ACID.
CC       IT IS A COMPONENT OF THE FATTY ACID REDUCTASE COMPLEX RESPONSIBLE
CC       FOR CONVERTING TETRADECANOIC ACID TO THE ALDEHYDE WHICH SERVES AS
CC       SUBSTRATE IN THE LUCIFERASE-CATALYZED REACTION.
CC   -!- CATALYTIC ACTIVITY: ATP + AN ACID + PROTEIN = AMP + PYROPHOSPHATE
CC       + AN ACYL-PROTEIN THIOLESTER.
CC   -!- PATHWAY: SECOND STEP IN THE BIOLUMINESCENT FATTY ACID REDUCTASE
CC       COMPLEX.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M62812; -; NOT_ANNOTATED_CDS.
KW   Luminescence; Ligase.
FT   NON_TER       1      1
SQ   SEQUENCE   3 AA;  374 MW;  6AA3303000000000 CRC64;

  Query Match             21.7%;  Score 10;  DB 1;  Length 3;
  Best Local Similarity   33.3%;  Pred. No. 8.8e+04;
  Matches    1;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        4 VQD 6
            ::|
Db        1 IKD 3


RESULT  26
TMOF_SARBU
ID   TMOF_SARBU     STANDARD;      PRT;     6 AA.
AC   P41495;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   TRYPSIN-MODULATING OOSTATIC FACTOR (TMOF).
OS   Sarcophaga bullata (Grey flesh fly) (Neobellieria bullata).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Sarcophagidae; Sarcophaga.
RN   [1]
RP   SEQUENCE, AND SYNTHESIS.
RC   TISSUE=OVARY;
RX   MEDLINE; 94211930.
RA   Bylemans D., Borovsky D., Hunt D.F., Shabanowitz J., Grauwels L.,
RA   de Loof A.;
RT   "Sequencing and characterization of trypsin modulating oostatic
RT   factor (TMOF) from the ovaries of the grey fleshfly, Neobellieria
RT   (Sarcophaga) bullata.";
RL   Regul. Pept. 50:61-72(1994).
CC   -!- FUNCTION: HAS AN OOSTATIC ACTIVITY. INHIBITS TRYPSIN BIOSYNTHESIS
CC       IN THE MIDGUT WHICH INDIRECTLY REDUCES THE VITELLOGENIN
CC       CONCENTRATION IN THE HEMOLYMPH RESULTING IN INHIBITION OF OOCYTE
CC       DEVELOPMENT.
CC   -!- DEVELOPMENTAL STAGE: SYNTHESIZED AND RELEASED FROM FOLLICULAR
```

```
CC       EPITHELIUM AFTER A BLOOD MEAL.
KW   Hormone.
SQ   SEQUENCE   6 AA;  695 MW;  61E72451B7642000 CRC64;


  Query Match             21.7%;  Score 10;  DB 1;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 NL 9
          ||
Db      4 NL 5


RESULT  27
TRPI_PSEPU
ID   TRPI_PSEPU     STANDARD;      PRT;     6 AA.
AC   P36414;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   TRPBA OPERON TRANSCRIPTIONAL ACTIVATOR (FRAGMENT).
GN   TRPI.
OS   Pseudomonas putida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PPG1 C1S;
RX   MEDLINE; 89335826.
RA   Eberly L., Crawford I.P.;
RT   "DNA sequence of the tryptophan synthase genes of Pseudomonas
RT   putida.";
RL   Biochimie 71:521-531(1989).
CC   -!- FUNCTION: ACTIVATES THE EXPRESSION OF THE TRPBA GENES ENCODING
CC       THE TWO TRYPTOPHAN SYNTHASE SUBUNITS. IN THE ABSENCE OF THE
CC       INDUCER (INDOLEGLYCEROL PHOSPHATE), TRPI BINDS UPSTREAM OF THE
CC       TRPAB OPERON, OVERLAPPING ITS OWN PROMOTER REGION.
CC   -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC       REGULATORS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X13299; CAA31660.1; -.
DR   INTERPRO; IPR000847; -.
DR   PROSITE; PS00044; HTH_LYSR_FAMILY; PARTIAL.
KW   Tryptophan biosynthesis; Transcription regulation; Activator;
KW   DNA-binding.
FT   NON_TER       6      6
SQ   SEQUENCE   6 AA;  683 MW;  77672AA1EDD6F000 CRC64;


  Query Match             21.7%;  Score 10;  DB 1;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 DL 7
          ||
Db      4 DL 5


RESULT  28
ALL7_CYDPO
ID   ALL7_CYDPO     STANDARD;      PRT;     7 AA.
AC   P82158;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYDIASTATIN 7.
OS   Cydia pomonella (Codling moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Tortricoidea; Tortricidae; Olethreutinae; Cydia.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LARVA;
RX   MEDLINE; 98054539.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
RA   Davey M., East P.D., Thorpe A.;
RT   "Lepidopteran peptides of the allatostatin superfamily.";
RL   Peptides 18:1301-1309(1997).
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  873 MW;  672879CABB569350 CRC64;


  Query Match             21.7%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      6 DLNL 9
          |  |
Db      4 DFGL 7


RESULT  29
FAR2_PROCL
ID   FAR2_PROCL     STANDARD;      PRT;     7 AA.
AC   P38498;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CARDIOEXCITATORY FMRFAMIDE HOMOLOG DF2.
OS   Procambarus clarkii (Red swamp crayfish).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC   Astacoidea; Cambaridae; Procambarus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PERICARDIAL ORGANS;
RX   MEDLINE; 93248032.
RA   Mercier A.J., Orchard I., Tebrugge V., Skerrett M.;
RT   "Isolation of two FMRFamide-related peptides from crayfish
RT   pericardial organs.";
RL   Peptides 14:137-143(1993).
CC   -!- FUNCTION: INCREASES THE RATE AND AMPLITUDE OF SPONTANEOUS
CC       CONTRACTIONS OF SEMI-ISOLATED HEARTS. INCREASES THE AMPLITUDE OF
CC       EXCITATORY POSTSYNAPTIC POTENTIALS IN ABDOMINAL EXTENSOR MUSCLE.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  967 MW;  69D40729C4540AC0 CRC64;


  Query Match             21.7%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      6 DLN 8
          | |
Db      1 DRN 3


RESULT  30
UN06_PINPS
ID   UN06_PINPS     STANDARD;      PRT;     7 AA.
AC   P81675;
DT   15-JUL-1999 (Rel. 38, Created)
```

```
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE OF NEEDLES (N141) (FRAGMENT).
OS   Pinus pinaster (Maritime pine).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Coniferopsida; Coniferales; Pinaceae; Pinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=NEEDLE;
RX   MEDLINE; 99274088.
RA   Costa P., Pionneau C., Bauw G., Dubos C., Bahrman N., Kremer A.,
RA   Frigerio J.-M., Plomion C.;
RT   "Separation and characterization of needle and xylem maritime pine
RT   proteins.";
RL   Electrophoresis 20:1098-1108(1999).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 6.6, ITS MW IS: 25 KDA.
FT   NON_TER       1      1
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  823 MW;  69D76724486B5740 CRC64;

  Query Match             21.7%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLNV 4
           | |:
Db       2 YGNL 5


RESULT  31
ACT_CARMA
ID   ACT_CARMA      STANDARD;      PRT;     8 AA.
AC   P80709;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   ACTIN (FRAGMENT).
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RA   Lachaise F., Somme G., Carpentier G., Granjeon E., Webster S.,
RA   Baghdassarian D.;
RT   "A transaldolase. An enzyme implicated in crab steroidogenesis.";
RL   Endocrine 5:23-32(1996).
CC   -!- FUNCTION: ACTINS ARE HIGHLY CONSERVED PROTEINS THAT ARE INVOLVED
CC       IN VARIOUS TYPES OF CELL MOTILITY AND ARE UBIQUITOUSLY EXPRESSED
CC       IN ALL EUKARYOTIC CELLS.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS PROTEIN IS:
CC       6.8, ITS MW IS: 46 KDA.
CC   -!- SIMILARITY: BELONGS TO THE ACTIN FAMILY.
DR   INTERPRO; IPR000279; -.
DR   PROSITE; PS00406; ACTINS_1; PARTIAL.
DR   PROSITE; PS00432; ACTINS_2; PARTIAL.
DR   PROSITE; PS01132; ACTINS_ACT_LIKE; PARTIAL.
KW   Structural protein.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  976 MW;  1424005AB2CAAEB3 CRC64;

  Query Match             21.7%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   25.0%;  Pred. No. 8.8e+04;
  Matches    1;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DLNL 9
           |:::
Db       3 DVDI 6
```

```
RESULT  32
AKH_MELML
ID   AKH_MELML      STANDARD;      PRT;     8 AA.
AC   P25423;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   ADIPOKINETIC HORMONE (AKH).
OS   Melolontha melolontha (Cockchafer),
OS   Geotrupes stercorosus (Dor beetle), and
OS   Pachnoda marginata (Flower beetle).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC   Scarabaeiformia; Scarabaeidae; Melolonthinae; Melolontha.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=M.MELOLONTHA, AND G.STERCOROSUS; TISSUE=CORPORA CARDIACA;
RX   MEDLINE; 91248100.
RA   Gaede G.;
RT   "A unique charged tyrosine-containing member of the adipokinetic
RT   hormone/red-pigment-concentrating hormone peptide family isolated and
RT   sequenced from two beetle species.";
RL   Biochem. J. 275:671-677(1991).
RN   [2]
RP   SEQUENCE.
RC   SPECIES=P.MARGINATA; TISSUE=CORPORA CARDIACA;
RX   MEDLINE; 92265187.
RA   Gaede G., Lopata A., Kellner R., Rinehart K.L. Jr.;
RT   "Primary structures of neuropeptides isolated from the corpora
RT   cardiaca of various cetonid beetle species determined by
RT   pulsed-liquid phase sequencing and tandem fast atom bombardment mass
RT   spectrometry.";
RL   Biol. Chem. Hoppe-Seyler 373:133-142(1992).
CC   -!- FUNCTION: THIS HORMONE, RELEASED FROM CELLS IN THE CORPORA
CC       CARDIACA AFTER THE BEGINNING OF FLIGHT, CAUSES RELEASE OF
CC       DIGLYCERIDES FROM THE FAT BODY AND THEN STIMULATES THE FLIGHT
CC       MUSCLES TO USE THESE DIGLYCERIDES AS AN ENERGY SOURCE.
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; S15422; S15422.
DR   PIR; S21663; S21663.
DR   INTERPRO; IPR002047; -.
DR   PROSITE; PS00256; AKH; 1.
KW   Neuropeptide; Amidation; Flight.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  1022 MW;  867AB775AB544736 CRC64;

  Query Match             21.7%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LN 3
           ||
Db       2 LN 3


RESULT  33
ALL5_CYDPO
ID   ALL5_CYDPO     STANDARD;      PRT;     8 AA.
AC   P82156;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYDIASTATIN 5.
OS   Cydia pomonella (Codling moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Tortricoidea; Tortricidae; Olethreutinae; Cydia.
RN   [1]
RP   SEQUENCE.
```

```
RC   TISSUE=LARVA;
RX   MEDLINE; 98054539.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
RA   Davey M., East P.D., Thorpe A.;
RT   "Lepidopteran peptides of the allatostatin superfamily.";
RL   Peptides 18:1301-1309(1997).
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  898 MW;  922879CABB58640D CRC64;


  Query Match             21.7%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 DLNL 9
           |  |
Db       5 DFGL 8


RESULT  34
ANG2_BOTJA
ID   ANG2_BOTJA     STANDARD;      PRT;     8 AA.
AC   Q10582;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   ANGIOTENSIN-LIKE PEPTIDE II (FRAGMENT).
OS   Bothrops jararaca (Jararaca).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Viperidae; Crotalinae; Bothrops.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PLASMA;
RX   MEDLINE; 96208932.
RA   Borgheresi R.A.M.B., Dalle Lucca J., Carmona E., Picarelli Z.P.;
RT   "Isolation and identification of angiotensin-like peptides from the
RT   plasma of the snake Bothrops jararaca.";
RL   Comp. Biochem. Physiol. 113B:467-473(1996).
CC   -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR   INTERPRO; IPR000215; -.
DR   PROSITE; PS00284; SERPIN; PARTIAL.
KW   Vasoconstrictor; Plasma; Serpin.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  1046 MW;  DDD761E04B42D40A CRC64;


  Query Match             21.7%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   33.3%;  Pred. No. 8.8e+04;
  Matches    1;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLN 3
           |::
Db       4 YIH 6


RESULT  35
B44K_PORGI
ID   B44K_PORGI     STANDARD;      PRT;     8 AA.
AC   P81886;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   44 KDA IMMUNOGENIC PROTEIN (FRAGMENT).
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; CFB group; Bacteroidaceae; Porphyromonas.
RN   [1]
RP   SEQUENCE.
RC   STRAIN=VPB 3492;
RX   MEDLINE; 20198497.
RA   Norris J.M., Love D.N.;
RT   "Serum antibody responses of cats to soluble whole cell antigens of
RT   feline porphyromonas gingivalis.";
RL   Vet. Microbiol. 73:37-49(2000).
CC   -!- SIMILARITY: TO P.GINGIVALIS HEMAGGLUTININ A.
KW   Antigen.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  989 MW;  9554540326CB476D CRC64;

  Query Match             21.7%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 QDLNL 9
           |  |:
Db       4 QKRNI 8


Search completed: December 16, 2000, 04:23:32
Job time: 4567 sec
```

```
                        GenCore version 4.5
               Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 03:07:19 ; Search time 107.12 Seconds
                                               (without alignments)
                                               2.873 Million cell updates/sec

Title:           US-09-529-121-9
Perfect score:   46
Sequence:        1 YLNVQDLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters:      61695

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       A_Geneseq_36:*
                1:  /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
                2:  /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
                3:  /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
                4:  /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
                5:  /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
                6:  /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
                7:  /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
                8:  /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
                9:  /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
                10:  /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
                11:  /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
                12:  /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
                13:  /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
                14:  /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
                15:  /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
                16:  /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
                17:  /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
                18:  /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
                19:  /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
                20:  /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
                21:  /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                %
Result         Query
  No.   Score  Match Length DB  ID                     Description
  -----------------------------------------------------------------------
    1     28    60.9      9  20  Y09526                 Carcinoembryonic a
    2     26    56.5      9  20  Y09527                 Carcinoembryonic a
    3     24    52.2      9  20  Y47685                 Immunogenic peptid
    4     24    52.2      9  20  Y47858                 Immunogenic peptid
    5     23    50.0      5  18  W45373                 Peptide antagonist
    6     23    50.0      8  17  W12586                 SH2 binding peptid
    7     23    50.0      8  20  Y25497                 Insulin-like growt
    8     23    50.0      9  11  R06899                 Human elastase inh
    9     23    50.0      9  16  R69827                 Peptide comprising
   10     23    50.0      9  16  R64152                 Elastase inhibitor
   11     23    50.0      9  18  W39723                 Human carcina-embr
   12     23    50.0      9  19  W77134                 CEA synthetic pept
   13     23    50.0      9  19  W70045                 CEA derived HLA-A2
   14     23    50.0      9  19  W46895                 SHC phosphopeptide
   15     23    50.0      9  20  Y48656                 Membrane dipeptida
   16     23    50.0      9  20  Y47655                 Immunogenic peptid
   17     23    50.0      9  20  Y25242                 HIV pol protein de
   18     23    50.0      9  20  Y09525                 Carcinoembryonic a
   19     22    47.8      6  14  R42432                 Sequence used to d
   20     22    47.8      9  20  Y47532                 Immunogenic peptid
   21     22    47.8      9  21  Y98547                 WT1 derived immuno
   22     22    47.8      9  21  Y98586                 WT1 derived immuno
   23     22    47.8      9  21  Y98691                 WT1 derived immuno
   24     22    47.8      9  21  Y98749                 WT1 derived immuno
   25     22    47.8      9  21  Y98783                 WT1 derived immuno
   26     21    45.7      6  19  W61747                 Peptide mimetic of
   27     21    45.7      7  21  Y63740                 Desmoglein cell ad
   28     21    45.7      8  19  W77008                 Peptide mimetic of
   29     21    45.7      8  21  Y63743                 Desmoglein cell ad
   30     21    45.7      9  15  R59112                 Peptide fragment (
   31     21    45.7      9  15  R59233                 Peptide fragment (
   32     21    45.7      9  16  R70596                 HIV(B35)GAG-8, hum
   33     21    45.7      9  16  R70067                 Control peptide 92
   34     21    45.7      9  17  Y07297                 Peptide against SH
   35     21    45.7      9  19  W78574                 SH2 domain binding
   36     21    45.7      9  19  W78670                 SH2 domain binding
   37     21    45.7      9  19  W78663                 SH2 domain binding
   38     21    45.7      9  19  W54515                 Synthetic polypept
   39     21    45.7      9  20  Y09528                 Carcinoembryonic a
   40     21    45.7      9  20  Y10265                 T cell epitope/MHC
   41     21    45.7      9  21  Y63746                 Desmoglein cell ad
   42     20    43.5      6  13  R29385                 Endothelin antagon
   43     20    43.5      6  15  R69163                 Endothelin C-termi
   44     20    43.5      8  20  Y38450                 Human secreted pro
   45     20    43.5      9  15  Y38184                 Hepatitis C virus-
   46     20    43.5      9  15  R49218                 HLA-A1 HCV antigen
   47     20    43.5      9  17  W49674                 Human leucocyte an
   48     20    43.5      9  20  Y45752                 Immunogenic peptid
   49     20    43.5      9  21  Y85450                 Tumour antigen pro
   50     20    43.5      9  21  Y54173                 HLA binding peptid
   51     20    43.5      9  21  Y59505                 Human NEDD8 protei
   52     19    41.3      4  18  W08875                 4-Aminobenzoyl-Tyr
   53     19    41.3      5  17  W00441                 Interleukin-6 anta
   54     19    41.3      5  18  W45368                 Peptide antagonist
   55     19    41.3      5  18  W45375                 Peptide antagonist
   56     19    41.3      5  18  W45376                 Peptide antagonist
   57     19    41.3      6   9  P82514                Viral inhibiting p
   58     19    41.3      6  12  R14520                 MCF peptide #32.
   59     19    41.3      7  18  W33711                 B. thuringiensis t
   60     19    41.3      8  16  R69833                 Peptide comprising
   61     19    41.3      9  15  R73765                 Antigen fragment 8
   62     19    41.3      9  16  R83937                 MHC class I restri
   63     19    41.3      9  16  R78852                 HIV gp41 586-594 c
   64     19    41.3      9  16  R69828                 Peptide comprising
   65     19    41.3      9  17  W12595                 SH2 binding peptid
   66     19    41.3      9  17  W12600                 SH2 binding peptid
   67     19    41.3      9  17  W12601                 SH2 binding peptid
   68     19    41.3      9  17  W12560                 SH2 binding peptid
   69     19    41.3      9  18  W22037                 Antigenic P60 pept
   70     19    41.3      9  20  Y46716                 Immunogenic peptid
   71     19    41.3      9  20  Y47492                 Immunogenic peptid
   72     19    41.3      9  20  Y09529                 Carcinoembryonic a
   73     18    39.1      4  17  W02180                 Residues 177-180 o
   74     18    39.1      4  18  W45369                 Peptide antagonist
   75     18    39.1      5  17  W02182                 Residues 176-180 o

                              ALIGNMENTS


RESULT   1
Y09526
ID   Y09526 standard; peptide; 9 AA.
XX
AC   Y09526;
```

```
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:2.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;


  Query Match             60.9%;  Score 28;  DB 20;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLNL 9
            ||:  ||||
Db        1 ylsgadlnl 9

RESULT   2
Y09527
ID   Y09527 standard; peptide; 9 AA.
XX
AC   Y09527;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:3.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
```

```
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;


  Query Match             56.5%;  Score 26;  DB 20;  Length 9;
  Best Local Similarity   55.6%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLNL 9
            ||:  |:||
Db        1 ylsgadinl 9

RESULT   3
Y47685
ID   Y47685 standard; Peptide; 9 AA.
XX
AC   Y47685;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2296.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
```

```
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 119; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;


  Query Match             52.2%;  Score 24;  DB 20;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLNVQDL 7
           ||::| |
Db       3 ylhiqsl 9


RESULT   4
Y47858
ID   Y47858 standard; Peptide; 9 AA.
XX
AC   Y47858;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2469.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
```

```
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 125; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;


  Query Match             52.2%;  Score 24;  DB 20;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLNVQDL 7
           ||::| |
Db       3 ylhiqsl 9


RESULT   5
W45373
ID   W45373 standard; peptide; 5 AA.
XX
AC   W45373;
XX
DT   11-MAY-1998  (first entry)
XX
DE   Peptide antagonist (FCE 28883) of cellular mitogenesis and motogenesis.
XX
KW   Peptide antagonist; mitogenesis; motogenesis; proliferative disease;
KW   cancer; tumour; phosphotyrosine; motif; SH2 domain.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "N-terminal acetyl-Tyr(PO3H2)"
FT   Modified-site   5
FT                   /note= "C-terminal amide"
XX
PN   WO9730079-A1.
XX
PD   21-AUG-1997.
XX
PF   10-FEB-1997;   97WO-EP00595.
XX
PR   15-FEB-1996;   96GB-0003227.
XX
```

```
PA   (PHAA ) PHARMACIA & UPJOHN SPA.
XX
PI   Bardelli A,  Battistini C,  Comoglio P,  Corradi F;
PI   De Rosa S,  Giordano P;
XX
DR   WPI; 1997-424973/39.
XX
PT   New peptide with phosphorylated tyrosine residue - is useful for
PT   treating e.g. proliferative, inflammatory, allergic, auto:immune,
PT   viral and cardiovascular diseases
XX
PS   Claim 11; Page 72; 84pp; English.
XX
CC   The present sequence represents a specifically claimed peptide with
CC   a phosphorylated tyrosine residue, which can bind to a cytosolic signal
CC   transducer and inhibit binding to a receptor tyrosine kinase or to
CC   another tyrosine phosphorylated transducer. It can interfere with cell
CC   motility, cell proliferation, cell invasiveness and tubulogenosis. It can
CC   be used in the control or treatment of proliferative diseases such as
CC   cancer, tumour growth and/or metastasis, and psoriasis, and in the
CC   control or treatment of inflammatory, allergic, autoimmune, viral and
CC   cardiovascular diseases.
XX
SQ   Sequence   5 AA;


  Query Match             50.0%;  Score 23;  DB 18;  Length 5;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLNVQ 5
           |:|||
Db       1 yvnvq 5


RESULT   6
W12586
ID   W12586 standard; peptide; 8 AA.
XX
AC   W12586;
XX
DT   08-APR-1997  (first entry)
XX
DE   SH2 binding peptide #16.
XX
KW   Core peptide; src homology region 2 binding peptide; SH2; cell growth;
KW   differentiation; regulation; receptor tyrosine kinase pathway; cancer;
KW   signal transduction pathway; non-insulin dependent diabetes;
KW   insulin-resistant diabetes.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /label= OTHER
FT                   /note= "opt. selected from tyrosine, phosphotyrosine,
FT                          tyrosine isosteres, and phosphotyrosine
FT                          isosteres"
FT   Modified-site   5
FT                   /label= OTHER
FT                   /note= "opt. selected from tyrosine, phosphotyrosine,
FT                          tyrosine isosteres, and phosphotyrosine
FT                          isosteres"
XX
PN   WO9623813-A1.
XX
PD   08-AUG-1996.
XX
PF   31-JAN-1996;   96WO-US01544.
XX
PR   01-FEB-1995;   95US-0382100.
XX
PA   (AFFY-) AFFYMAX TECHNOLOGIES NV.
```

```
XX
PI   Gordeev MF,  Gordon E,  Grove JR,  Hart CP,  Kim MH;
PI   Patel DV,  Szardenings AK;
XX
DR   WPI; 1996-371373/37.
XX
PT   Peptide(s) which bind to SH2 domains - are used to treat diseases
PT   associated with aberrant cell growth, differentiation or regulation
PT   associated with defects in receptor tyrosine kinase pathways
XX
PS   Claim 14; Page 118; 203pp; English.
XX
CC   The sequences given in W12571-614 represent src homology region 2 (SH2)
CC   binding peptides which contain a core peptide which corresponds to the
CC   formula:
CC                        Z7-X-Z8-X
CC   X = any D- or L- amino acid;
CC   Z7 = phosphotyrosine or its isostere;
CC   Z8 = asparagine or its isostere;
CC   the amino terminus is acylated, and the peptide is less than 14
CC   residues in length, with the proviso that if Z7 is phosphotyrosine
CC   and Z8 is asparagine, then the peptide is not GDGZ7XZ8XPLLL.
CC   These SH2 binding peptides are used to treat of diagnose diseases
CC   associated with aberrant cell growth, differentiation or regulation
CC   which is associated with defects in receptor tyrosine kinase pathways,
CC   by partially blocking or inhibiting a cellular signal transduction
CC   pathway.  The disease may be cancer, a developmental or differentiation
CC   disease or insulin-resistant (or non-insulin dependent) diabetes.
XX
SQ   Sequence   8 AA;


  Query Match             50.0%;  Score 23;  DB 17;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLNVQD 6
           |:|| |
Db       1 yvnvyd 6


RESULT   7
Y25497
ID   Y25497 standard; peptide; 8 AA.
XX
AC   Y25497;
XX
DT   22-SEP-1999  (first entry)
XX
DE   Insulin-like growth factor binding protein fragment #32.
XX
KW   Insulin-like growth factor binding protein; IBP; immunoprotective;
KW   antiproliferative; anti-inflammatory; cytostatic; treatment; muscle loss;
KW   shrinkage; wastage; osteoporosis; diabetes; amyloid lateral sclerosis;
KW   neuropathy; inflammation; cancer; neoplastic disease; growth disturbance;
KW   muscle; bone apparatus; wound healing; bone healing; gene therapy;
KW   diagnosis; nervous system; lymph gland; stomach; intestinal tract.
XX
OS   Unidentified.
XX
PN   DE19757250-A1.
XX
PD   01-JUL-1999.
XX
PF   22-DEC-1997;   97DE-1057250.
XX
PR   22-DEC-1997;   97DE-1057250.
XX
PA   (FORS/) FORSSMANN W.
XX
PI   Forssmann W,  Kling L,  Mostafavi H,  Obendorf M,  Opitz H;
PI   Staendker L;
XX
```

```
DR   WPI; 1999-372233/32.
XX
PT   An insulin-like growth factor binding protein
XX
PS   Claim 10; Page 18; 26pp; German.
XX
CC   This invention describes a novel insulin-like growth factor binding
CC   protein (IBP) constructed using the formula R1-C-X1-PNC-X2-QC-X3-CWCV-
CC   X4-C-R2 where cyclic, glycosylated, phosphorylated, acetylated, amidated,
CC   sulphated derivatives and fragments of these with the physiological
CC   ability of IBP are included and R1NH2 = an amino acid or peptide with a
CC   sequence of up to 41 amino acid residues, X1 = a peptide consisting of
CC   24-31 amino acids, X2 = a peptide consisting of 9 amino acid, X3 = a
CC   peptide consisting of 10 amino acids, X4 = a peptide consisting of 18-24
CC   amino acids, R2COOH = CONH2 or a peptide with up to 12 amino acids. The
CC   products of the invention have immunoprotective, antiproliferative,
CC   anti-inflammatory and cytostatic activity. The IBP peptides, complexes,
CC   antibodies, inhibitors and (antisense) nucleic acids are all useful in
CC   the treatment of muscle loss/shrinkage/wastage, osteoporosis, diabetes,
CC   amyloid lateral sclerosis, peripheral and central neuropathies,
CC   inflammatory processes, cancer, inflammatory and neoplastic diseases,
CC   growth disturbances, diseases of the muscles, diseases of the bone
CC   apparatus and wound or bone healing. The nucleic acids and antisense
CC   oligonucleotides can be used in somatic or non-somatic gene therapy. The
CC   peptides, etc., can be used in diagnosis of functional deficiencies in
CC   bones, muscles, nervous systems, lymph glands, stomach/intestinal tracts,
CC   immune systems, for diabetes, inflammatory and neoplastic processes and
CC   also as a marker for cancer. This sequence represents a specific claimed
CC   example of a peptide which can occupy the R2 position of the IBP of the
CC   invention.
XX
SQ   Sequence   8 AA;


  Query Match             50.0%;  Score 23;  DB 20;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLNVQD 6
           | |||:
Db       3 yfnvqn 8


RESULT   8
R06899
ID   R06899 standard; peptide; 9 AA.
XX
AC   R06899;
XX
DT   17-JAN-1991  (first entry)
XX
DE   Human elastase inhibitor peptide.
XX
KW   Elastase; inflammation; genetic diseases; neoplasm; tuberculosis.
XX
OS   Homo sapiens.
XX
PN   WO9009737-A.
XX
PD   07-SEP-1990.
XX
PF   22-FEB-1990;   90WO-US00920.
XX
PR   23-FEB-1989;   89US-0314383.
XX
PA   (BLOO-) CENT BLOOD RES.
XX
PI   Remold-O'Donnell E;
XX
DR   WPI; 1990-290106/38.
XX
PT   New purified human elastase inhibitor - with therapeutic and
PT   diagnostic applications
```

```
XX
PS   Claim 6; page 34; 49pp; English.
XX
CC   This is a constituent peptide of a purified human elastase in-
CC   hibitor (EI) which is capable of forming a covalent complex with
CC   elastase and acting as an inhibitor of the elastinolytic activity
CC   of elastase.  The EI is useful in the treatment of conditions
CC   such as cystic fibrosis or bronchiectasis.  The Abs may be used
CC   e.g. to diagnose rheumatological diseases, idiopathic diseases and
CC   inflammatory diseases or as a diagnostic tool for e.g. neoplastic
CC   or genetic diseases.  The corresp. gene may be used in gene
CC   replacement therapy.  See also R06900-08
XX
SQ   Sequence   9 AA;


  Query Match             50.0%;  Score 23;  DB 11;  Length 9;
  Best Local Similarity   83.3%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LNVQDL 7
           | ||||
Db       1 lgvqdl 6


RESULT   9
R69827
ID   R69827 standard; Peptide; 9 AA.
XX
AC   R69827;
XX
DT   20-SEP-1995  (first entry)
XX
DE   Peptide comprising a part of hHGFR intracellular region.
XX
KW   intracellular region; human hepatocyte growth factor; hHGF;
KW   receptor; cytosolic signal transducer binding; cell proliferation;
KW   movement; extracelllular matrix invasion.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "H-Asp"
FT   Modified-site   5
FT                   /note= "optionally phosphorylated"
FT   Modified-site   9
FT                   /note= "Gln-OH"
XX
PN   WO9501376-A.
XX
PD   12-JAN-1995.
XX
PF   15-JUN-1994;   94WO-EP01943.
XX
PR   30-JUN-1993;   93GB-0013528.
PR   18-APR-1994;   94GB-0007673.
XX
PA   (FARM ) FARMITALIA ERBA SRL CARLO.
XX
PI   Comoglio P,  Ponzetto C;
XX
DR   WPI; 1995-060954/08.
XX
PT   New peptide(s) which bind to a cytosolic signal transducer - used
PT   for interfering with pathways leading to cell proliferation and
PT   motility, for treating neoplastic disease
XX
PS   Claim 10; Page 64; 19pp; English.
XX
CC   This peptide is a tyrosine-contg. mol. representing a site of tyrosine
CC   phosphorylation. The peptides of the invention comprise a portion of
CC   the intracellular region of the human hepatocyte growth factor receptor
```

```
CC   (hHGFR), and can generally reproduce potential recognition motifs for
CC   the SH2 domains of intracellular (cytosolic) signal inducers. The
CC   peptides can interfere with pathways leading to cell proliferation,
CC   movement and extracellular matrix invasion. They can be used to inhibit
CC   growth of neoplastic cells and to prevent metastatic spreading.
XX
SQ   Sequence   9 AA;

  Query Match             50.0%;  Score 23;  DB 16;  Length 9;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLNVQ 5
              |:|||
Db          5 yvnvq 9

RESULT  10
R64152
ID   R64152 standard; Peptide; 9 AA.
XX
AC   R64152;
XX
DT   07-AUG-1995  (first entry)
XX
DE   Elastase inhibitor peptide.
XX
KW   Elastase inhibitor; monocyte; probe; hybridization.
XX
OS   Homo sapiens.
XX
PN   US5370991-A.
XX
PD   06-DEC-1994.
XX
PF   23-FEB-1989;   89US-0314383.
XX
PR   23-FEB-1989;   89US-0314383.
PR   06-SEP-1991;   91US-0755461.
XX
PA   (BLOO-) CENT BLOOD RES INC.
XX
PI   Remold-O'Donnell E;
XX
DR   WPI; 1995-021883/03.
XX
PT   DNA encoding human monocyte elastase inhibitor - used to develop
PT   prods. for diagnosis and therapy of conditions involving elastase
PT   or elastase inhibitor
XX
PS   Claim 1; Column 25-26; 22pp; English.
XX
CC   The peptides given in R64148-58 are tryptic digests of human
CC   monocyte elastase inhibitor (hMEI) (R64159), and were used to
CC   design oligonucleotide probes for the isolation of the hMEI
CC   gene (Q75377).
XX
SQ   Sequence   9 AA;

  Query Match             50.0%;  Score 23;  DB 16;  Length 9;
  Best Local Similarity   83.3%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LNVQDL 7
              | ||||
Db          1 lgvqdl 6

RESULT  11
W39723
ID   W39723 standard; peptide; 9 AA.
XX
AC   W39723;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Human carcina-embryonic antigen (CEA) peptide (pos. 571-579).
XX
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
XX
OS   Homo sapiens.
XX
PN   WO9741440-A1.
XX
PD   06-NOV-1997.
XX
PF   28-APR-1997;   97WO-NL00229.
XX
PR   23-DEC-1996;   96EP-0203670.
PR   26-APR-1996;   96EP-0201145.
XX
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
PI   Kast WM,  Melief CJM,  Offringa R,  Toes REM,  Van Der Burg SH;
XX
DR   WPI; 1997-549891/50.
XX
PT   Method of selecting T cell peptide epitope(s) - by measuring the
PT   stability of HLA class I-peptide complexes on intact B cells
XX
PS   Example 3; Page 85; 109pp; English.
XX
CC   Peptides W39430-W39734 are used in a novel method for the selection of
CC   immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC   method involves the identification of peptide sequences capable of
CC   binding to an HLA (human leukocyte antigen) class I molecule and
CC   measuring the binding of this epitope peptide to the HLA class I peptide.
CC   The stability of binding of the peptide and MHC (major histocompatablilty
CC   complex) class I molecule is measured on intact human B cells carrying
CC   the MHC molecule at their cell surfaces. The method can be used to select
CC   peptide epitopes for generating vaccines against a disease associated
CC   with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC   especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC   immune responses. Peptide W39723 is derived from the human
CC   carcino-embryonic antigen (CEA) and has the ability to bind to the human
CC   MHC Class I allele HLA-A2.1.
XX
SQ   Sequence   9 AA;

  Query Match             50.0%;  Score 23;  DB 18;  Length 9;
  Best Local Similarity   55.6%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLNVQDLNL 9
              ||:  :|||
Db          1 ylsganlnl 9

RESULT  12
W77134
ID   W77134 standard; peptide; 9 AA.
XX
AC   W77134;
XX
DT   16-NOV-1998  (first entry)
XX
DE   CEA synthetic peptide epitope 1.
XX
KW   Tyrosinase; tyrosinase cytotoxic lymphocyte response;
KW   cytotoxic T lymphocyte; cysteine-depleted; melanoma.
```

```
XX
OS   Synthetic.
XX
PN   WO9833810-A2.
XX
PD   06-AUG-1998.
XX
PF   29-JAN-1998;   98WO-US01592.
XX
PR   30-JAN-1997;   97US-0037781.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI   Engelhard VH,  Hunt DF,  Kittlesen D,  Slingluff CL;
XX
DR   WPI; 1998-437388/37.
XX
PT   Disease specific immunogen  - comprises disease specific cytotoxic T
PT   lymphocyte epitope used to elicit melanoma specific CTL response
XX
PS   Disclosure; Page 27; 93pp; English.
XX
CC   The peptide epitope W77119-W77138 were created for human tumour-specific
CC   cytotoxic T lymphocyte response.  These peptides are are cysteine-
CC   depleted mutants of a native disease-specific CTL epitope.  The cysteine-
CC   depleted CTL epitopes elicit a stronger or more specific CTL response
CC   than the native epitope.  The epitopes can be used in a disease-specific
CC   immunogen to protect a mammal against disease in particular melanomas.
CC   The peptides may also be used to screen a sample for the presence of
CC   an antigen with the same epitope, or with a different cross-reactive
CC   epitope.
XX
SQ   Sequence   9 AA;


  Query Match             50.0%;  Score 23;  DB 19;  Length 9;
  Best Local Similarity   55.6%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLNVQDLNL 9
                ||:  :|||
Db          1 ylsganlnl 9


RESULT  13
W70045
ID   W70045 standard; peptide; 9 AA.
XX
AC   W70045;
XX
DT   22-OCT-1998  (first entry)
XX
DE   CEA derived HLA-A2.1 binding peptide 2 (residues 605-613).
XX
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW   antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW   viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW   fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9833888-A1.
XX
PD   06-AUG-1998.
XX
PF   30-JAN-1998;   98WO-US01959.
XX
PR   31-JAN-1997;   97US-0036696.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
```

```
PI   Celis E,  Sette A,  Sidney J,  Southwood S,  Tsai V;
XX
DR   WPI; 1998-437445/37.
XX
PT   Production of antigen-specific cytotoxic T cells - by incubating
PT   immunogenic peptide(s) from antigen that binds class I major
PT   histocompatibility complex molecules with pre-treated antigen
PT   presenting cells
XX
PS   Example 6; Page 75; 104pp; English.
XX
CC   Sequences shown in W70044 to W70052 represent peptides derived from
CC   carcinoembryonic antigen (CEA). The peptides can bind to a human
CC   leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method
CC   of invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC   vitro. The method comprises contacting immunogenic peptides from an
CC   antigen that binds class I major histocompatibility complex (MHC)
CC   molecules with antigen presenting cells (APCs) pretreated with
CC   pretreatment growth factors, and incubating the APCs with purified CD8
CC   cells in the presence of at least 2 incubation growth factors, thereby
CC   producing antigen-specific CTLs. A method for specifically killing
CC   target cells in a human patient is also provided which comprises
CC   obtaining a fluid sample containing CTLs from a patient, contacting the
CC   cytotoxic T cells with APCs pretreated with pre-treatment growth factors,
CC   where the APCs comprise class I MHC molecules. The pretreated APCs are
CC   incubated with the cytotoxic growth factors, thereby producing activated
CC   CTLs which are contacted with a carrier to form a composition. The
CC   composition can then be administered to the patient. The activated CTLs
CC   can be used for treating cancers, immune disorders, viral infections,
CC   AIDS, hepatitis, bacterial infection, fungal infection, malaria or
CC   tuberculosis.
XX
SQ   Sequence   9 AA;


  Query Match             50.0%;  Score 23;  DB 19;  Length 9;
  Best Local Similarity   55.6%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLNVQDLNL 9
                ||:  :|||
Db          1 ylsganlnl 9


RESULT  14
W46895
ID   W46895 standard; peptide; 9 AA.
XX
AC   W46895;
XX
DT   19-JUN-1998  (first entry)
XX
DE   SHC phosphopeptide capable of binding to the SH2 domain of Grb2.
XX
KW   SHC phosphopeptide; binding; src homology 2 domain; SH2 domain; Grb2;
KW   signal transduction protein; non-phosphorylated; inhibition; treatment;
KW   hyper-proliferative disease; human cancer.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   5
FT                   /note= "phosphorylated"
XX
PN   WO9802176-A1.
XX
PD   22-JAN-1998.
XX
PF   16-JUL-1997;   97WO-US12501.
XX
PR   16-JUL-1996;   96US-0021858.
XX
PA   (GEOU ) UNIV GEORGETOWN.
```

```
PA   (UYVE-) UNIV VERMONT & STATE AGRIC COLLEGE.
XX
PI   King CR,  Krag D,  Oligino L,  Sastry L;
XX
DR   WPI; 1998-110340/10.
XX
PT   Non-phosphorylated peptide(s) that bind Src Homology 2 domain of
PT   signal transducing protein - at least as well as natural
PT   phosphorylated target, particularly from treatment of cancer
XX
PS   Disclosure; Page 9; 39pp; English.
XX
CC   The present sequence represents a SHC phosphopeptide that is capable of
CC   binding to the src homology 2 (SH2) domain of Grb2. Grb2 is a signal
CC   transduction protein. The specification describes a non-phosphorylated
CC   peptide (W46896) that binds to the SH2 domain of Grb2 with affinity
CC   similar to, or greater than, that of the present sequence. This
CC   non-phosphorylated peptide contains a tyrosine residue that has not been
CC   modified by phosphate or similar charged group. The non-phosphorylated
CC   peptide is used to inhibit a signal transduction process that involves
CC   binding of a phosphorylated protein or peptide to the SH2 domain
CC   of a signal transduction protein, particularly Grb2. It is used
CC   specifically for treatment of hyper-proliferative diseases, especially
CC   human cancer.
XX
SQ   Sequence   9 AA;


  Query Match             50.0%;  Score 23;  DB 19;  Length 9;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLNVQ 5
           |:|||
Db       5 yvnvq 9


RESULT  15
Y48656
ID   Y48656 standard; Peptide; 9 AA.
XX
AC   Y48656;
XX
DT   10-DEC-1999  (first entry)
XX
DE   Membrane dipeptidase-binding lung homing peptide #27.
XX
KW   Homing peptide; organ; tissue; lung; pancreas; skin; retina; MDP;
KW   prostate; ovary; lymph node; adrenal gland; liver; gut; tumour;
KW   membrane dipeptidase.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9946284-A2.
XX
PD   16-SEP-1999.
XX
PF   10-MAR-1999;   99WO-US05284.
XX
PR   13-MAR-1998;   98US-0042107.
PR   26-FEB-1999;   99US-0042107.
XX
PA   (BURN-) BURNHAM INST.
XX
PI   Rajotte D,  Pasqualini R,  Ruoslahti EI;
XX
DR   WPI; 1999-571717/48.
XX
PT   New peptides which selectively home to organs or tissues, used for,
PT   e.g. identifying target ligands and for therapy of pathological
PT   conditions  -
XX
```

```
PS   Example 6; Page 144; 193pp; English.
XX
CC   The present invention describes peptides that selectively home to a
CC   tissue or organ. The peptides can be used for identifying an organ
CC   or tissue, for identifying a target molecule expressed by an organ or
CC   tissue or for treating an organ or tissue pathology, where the organ or
CC   tissue is selected from prostate, lung, skin, retina, pancreas, gut,
CC   ovary, adrenal gland, liver, and lymph node. The peptide bind to the
CC   membrane dipeptidase (MDP). Y48618 to Y49066 represent sequences
CC   which are used in the exemplification of the present invention.
XX
SQ   Sequence   9 AA;


  Query Match             50.0%;  Score 23;  DB 20;  Length 9;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 QDLNL 9
           ||||:
Db       2 qdlni 6


RESULT  16
Y47655
ID   Y47655 standard; Peptide; 9 AA.
XX
AC   Y47655;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2266.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 118; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
```

```
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;


  Query Match             50.0%;  Score 23;  DB 20;  Length 9;
  Best Local Similarity   55.6%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLNL 9
            ||:  :|||
Db        1 ylsganlnl 9


RESULT  17
Y25242
ID   Y25242 standard; peptide; 9 AA.
XX
AC   Y25242;
XX
DT   03-SEP-1999  (first entry)
XX
DE   HIV pol protein derived peptide fragment #14.
XX
KW   Heat shock protein; HSP; complex; denatured protein matrix; antigen;
KW   vaccine; allergic disease; treatment; susceptibility; Th2; skin rash;
KW   allergic reaction; asthma; pol protein.
XX
OS   Human immunodeficiency virus.
XX
PN   WO9929182-A1.
XX
PD   17-JUN-1999.
XX
PF   04-DEC-1998;   98WO-US25734.
XX
PR   05-DEC-1997;   97US-0986234.
PR   05-DEC-1997;   97US-0985548.
XX
PA   (UYNE-) UNIV NEW MEXICO STATE.
XX
PI   Moseley PL,  Wallen ES;
XX
DR   WPI; 1999-394912/33.
XX
PT   Synthesizing heat shock protein complexes using a denatured protein
PT   matrix
XX
PS   Example 1; Fig 1C; 33pp; English.
XX
CC   This invention describes a novel method for synthesizing heat shock
CC   protein (HSP) complexes comprises adding a heat shock protein to a
CC   denatured protein matrix for binding, and adding a complexing solution
CC   comprising a peptide to elute a heat shock protein-peptide complex. A
CC   HSP-antigen complex is useful as a vaccine for treating an allergic
CC   disease (in a mammal, preferably a human) to reduce susceptibility of
CC   the Th2 response, the complex comprising a HSP-antigenic peptide complex.
CC   The complex is administered to prevent a mammal from having an allergic
CC   reaction to an allergic disease, or administered to a mammal having an
CC   allergic disease, to reduce the allergic reactions. Allergic diseases
CC   include asthma and skin rashes. Prior art methods or preventing/treating
CC   allergic diseases include antihistamines which treat only the symptoms,
CC   corticosteroids which have severe side effects and desensitization
CC   therapy which has limited uses. The new method also allows more
CC   flexibility of use of peptide-based vaccines, as prior art HSP-based
CC   vaccines require isolation from a portion of the tumour itself. This
CC   sequence represents a peptide fragment derived from the HIV pol protein
CC   which is used in the method of the invention.
XX
SQ   Sequence   9 AA;

  Query Match             50.0%;  Score 23;  DB 20;  Length 9;
  Best Local Similarity   55.6%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        1 YLNVQDLNL 9
            || :||  |
Db        1 ylalqdsgl 9


RESULT  18
Y09525
ID   Y09525 standard; peptide; 9 AA.
XX
AC   Y09525;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist CAP-1.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 1; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;


  Query Match             50.0%;  Score 23;  DB 20;  Length 9;
  Best Local Similarity   55.6%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy         1 YLNVQDLNL 9
             ||:  :|||
Db         1 ylsganlnl 9


RESULT  19
R42432
ID   R42432 standard; peptide; 6 AA.
XX
AC   R42432;
XX
DT   24-MAY-1994  (first entry)
XX
DE   Sequence used to design PCR primer for rat gustducin alpha subunit.
XX
KW   Taste modifying agent; ligand; antiligand; binding activity; taste;
KW   taste receptor cells; sweet; bitter; sweet; salty; sour;
KW   amplification.
XX
OS   Synthetic.
XX
PN   WO9321337-A.
XX
PD   28-OCT-1993.
XX
PF   08-APR-1993;   93WO-US03279.
XX
PR   09-APR-1992;   92US-0868353.
XX
PA   (MARG/) MARGOLSKEE R F.
XX
PI   Margolskee RF;
XX
DR   WPI; 1993-351746/44.
DR   N-PSDB; Q51100.
XX
PT   New gustducin alpha subunit protein - used for identifying taste
PT   modifying agents which mimic or inhibit sweet, bitter, salty or
PT   sour tastes
XX
PS   Disclosure; Page 10; 50pp; English.
XX
CC   Six degenerate oligonucleotide primer sets were made to correspond
CC   to regions of amino acids highly conserved among previously described
CC   G protein alpha subunits.  The primers were used for PCR with DNA from
CC   a taste cell library as template.  Partial clones and further PCR
CC   primers were used to isolate a composite gustatory alpha subunit
CC   clone or gustducin gene clone.
CC   See also R42405-36.
XX
SQ   Sequence   6 AA;

  Query Match             47.8%;  Score 22;  DB 14;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         1 YLNVQD 6
             :|| ||
Db         1 flnkqd 6


RESULT  20
Y47532
ID   Y47532 standard; Peptide; 9 AA.
XX
AC   Y47532;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2143.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
```

```
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 113; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;

  Query Match             47.8%;  Score 22;  DB 20;  Length 9;
  Best Local Similarity   55.6%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         1 YLNVQDLNL 9
             :|: | |||
Db         1 flskqylnl 9


RESULT  21
Y98547
ID   Y98547 standard; Peptide; 9 AA.
XX
AC   Y98547;
XX
DT   31-JUL-2000  (first entry)
XX
DE   WT1 derived immunogenic peptide SEQ ID NO:62.
XX
KW   WT1; immunotherapy; immunogenic; malignant disease; cancer; leukaemia;
KW   metastatic disease; mouse; human; Wilm's tumour; immune response;
```

```
KW   vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200018795-A2.
XX
PD   06-APR-2000.
XX
PF   30-SEP-1999;   99WO-US22819.
XX
PR   30-SEP-1998;   98US-0164223.
PR   25-MAR-1999;   99US-0276484.
XX
PA   (CORI-) CORIXA CORP.
PA   (GAIG/) GAIGER A.
XX
PI   Gaiger A,  Cheever M;
XX
DR   WPI; 2000-293107/25.
XX
PT   Novel polypeptides comprising an immunogenic portion of a native WT1
PT   polypeptide, useful for inhibiting the development of malignant
PT   diseases associated with WT1 expression e.g. leukemia or cancer  -
XX
PS   Claim 4; Page 153; 193pp; English.
XX
CC   The present invention describes polypeptides (I) comprising an
CC   immunogenic portion of a native Wilm's Tumour gene product polypeptide,
CC   WT1, (or variants of the immunogenic portion retaining the ability to
CC   react with WT1-specific antisera and/or T-cell lines or clones) and
CC   comprising 16 consecutive amino acids (aa) or less of a native WT1
CC   polypeptide. The polypeptides are useful therapeutically and to
CC   manufacture medicaments for enhancing/inducing an immune response in
CC   patients. The polypeptides, mimetics or polynucleotides can be included
CC   with a carrier/excipient in pharmaceutical compositions or with a
CC   non-specific immune response enhancer (e.g. an adjuvant or enhancer
CC   preferentially enhancing a T cell response) in vaccines. Pharmaceutical
CC   compositions and vaccines can be administered to human patients to
CC   enhance or induce an immune response specific for WT1 or a cell
CC   expressing WT1, useful to inhibit the development of malignant diseases
CC   associated with WT1 expression, e.g. leukemia (especially acute/chronic
CC   myeloid leukemia or acute lymphocytic leukemia) or cancer (especially
CC   breast, lung, thyroid or gastrointestinal cancer, or a melanoma). Y98501
CC   to Y98811 represent polypeptide sequences, and A13848 to A13862 represent
CC   PCR primers, used in the exemplification of the present invention.
XX
SQ   Sequence   9 AA;


  Query Match             47.8%;  Score 22;  DB 21;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 NVQDLN 8
            :|:|||
Db        1 dvrdln 6


RESULT  22
Y98586
ID   Y98586 standard; Peptide; 9 AA.
XX
AC   Y98586;
XX
DT   31-JUL-2000  (first entry)
XX
DE   WT1 derived immunogenic peptide SEQ ID NO:101.
XX
KW   WT1; immunotherapy; immunogenic; malignant disease; cancer; leukaemia;
KW   metastatic disease; mouse; human; Wilm's tumour; immune response;
KW   vaccine.
XX
OS   Homo sapiens.
```

```
XX
PN   WO200018795-A2.
XX
PD   06-APR-2000.
XX
PF   30-SEP-1999;   99WO-US22819.
XX
PR   30-SEP-1998;   98US-0164223.
PR   25-MAR-1999;   99US-0276484.
XX
PA   (CORI-) CORIXA CORP.
PA   (GAIG/) GAIGER A.
XX
PI   Gaiger A,  Cheever M;
XX
DR   WPI; 2000-293107/25.
XX
PT   Novel polypeptides comprising an immunogenic portion of a native WT1
PT   polypeptide, useful for inhibiting the development of malignant
PT   diseases associated with WT1 expression e.g. leukemia or cancer  -
XX
PS   Claim 4; Page 158; 193pp; English.
XX
CC   The present invention describes polypeptides (I) comprising an
CC   immunogenic portion of a native Wilm's Tumour gene product polypeptide,
CC   WT1, (or variants of the immunogenic portion retaining the ability to
CC   react with WT1-specific antisera and/or T-cell lines or clones) and
CC   comprising 16 consecutive amino acids (aa) or less of a native WT1
CC   polypeptide. The polypeptides are useful therapeutically and to
CC   manufacture medicaments for enhancing/inducing an immune response in
CC   patients. The polypeptides, mimetics or polynucleotides can be included
CC   with a carrier/excipient in pharmaceutical compositions or with a
CC   non-specific immune response enhancer (e.g. an adjuvant or enhancer
CC   preferentially enhancing a T cell response) in vaccines. Pharmaceutical
CC   compositions and vaccines can be administered to human patients to
CC   enhance or induce an immune response specific for WT1 or a cell
CC   expressing WT1, useful to inhibit the development of malignant diseases
CC   associated with WT1 expression, e.g. leukemia (especially acute/chronic
CC   myeloid leukemia or acute lymphocytic leukemia) or cancer (especially
CC   breast, lung, thyroid or gastrointestinal cancer, or a melanoma). Y98501
CC   to Y98811 represent polypeptide sequences, and A13848 to A13862 represent
CC   PCR primers, used in the exemplification of the present invention.
XX
SQ   Sequence   9 AA;


  Query Match             47.8%;  Score 22;  DB 21;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 NVQDLN 8
            :|:|||
Db        3 dvrdln 8


RESULT  23
Y98691
ID   Y98691 standard; Peptide; 9 AA.
XX
AC   Y98691;
XX
DT   31-JUL-2000  (first entry)
XX
DE   WT1 derived immunogenic peptide SEQ ID NO:206.
XX
KW   WT1; immunotherapy; immunogenic; malignant disease; cancer; leukaemia;
KW   metastatic disease; mouse; human; Wilm's tumour; immune response;
KW   vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200018795-A2.
XX
```

```
PD   06-APR-2000.
XX
PF   30-SEP-1999;   99WO-US22819.
XX
PR   30-SEP-1998;   98US-0164223.
PR   25-MAR-1999;   99US-0276484.
XX
PA   (CORI-) CORIXA CORP.
PA   (GAIG/) GAIGER A.
XX
PI   Gaiger A,  Cheever M;
XX
DR   WPI; 2000-293107/25.
XX
PT   Novel polypeptides comprising an immunogenic portion of a native WT1
PT   polypeptide, useful for inhibiting the development of malignant
PT   diseases associated with WT1 expression e.g. leukemia or cancer  -
XX
PS   Claim 4; Page 174; 193pp; English.
XX
CC   The present invention describes polypeptides (I) comprising an
CC   immunogenic portion of a native Wilm's Tumour gene product polypeptide,
CC   WT1, (or variants of the immunogenic portion retaining the ability to
CC   react with WT1-specific antisera and/or T-cell lines or clones) and
CC   comprising 16 consecutive amino acids (aa) or less of a native WT1
CC   polypeptide. The polypeptides are useful therapeutically and to
CC   manufacture medicaments for enhancing/inducing an immune response in
CC   patients. The polypeptides, mimetics or polynucleotides can be included
CC   with a carrier/excipient in pharmaceutical compositions or with a
CC   non-specific immune response enhancer (e.g. an adjuvant or enhancer
CC   preferentially enhancing a T cell response) in vaccines. Pharmaceutical
CC   compositions and vaccines can be administered to human patients to
CC   enhance or induce an immune response specific for WT1 or a cell
CC   expressing WT1, useful to inhibit the development of malignant diseases
CC   associated with WT1 expression, e.g. leukemia (especially acute/chronic
CC   myeloid leukemia or acute lymphocytic leukemia) or cancer (especially
CC   breast, lung, thyroid or gastrointestinal cancer, or a melanoma). Y98501
CC   to Y98811 represent polypeptide sequences, and A13848 to A13862 represent
CC   PCR primers, used in the exemplification of the present invention.
XX
SQ   Sequence   9 AA;


  Query Match             47.8%;  Score 22;  DB 21;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 NVQDLN 8
           :|:|||
Db       2 dvrdln 7


RESULT  24
Y98749
ID   Y98749 standard; Peptide; 9 AA.
XX
AC   Y98749;
XX
DT   31-JUL-2000  (first entry)
XX
DE   WT1 derived immunogenic peptide SEQ ID NO:264.
XX
KW   WT1; immunotherapy; immunogenic; malignant disease; cancer; leukaemia;
KW   metastatic disease; mouse; human; Wilm's tumour; immune response;
KW   vaccine.
XX
OS   Mus musculus.
XX
PN   WO200018795-A2.
XX
PD   06-APR-2000.
XX
PF   30-SEP-1999;   99WO-US22819.
```

```
XX
PR   30-SEP-1998;   98US-0164223.
PR   25-MAR-1999;   99US-0276484.
XX
PA   (CORI-) CORIXA CORP.
PA   (GAIG/) GAIGER A.
XX
PI   Gaiger A,  Cheever M;
XX
DR   WPI; 2000-293107/25.
XX
PT   Novel polypeptides comprising an immunogenic portion of a native WT1
PT   polypeptide, useful for inhibiting the development of malignant
PT   diseases associated with WT1 expression e.g. leukemia or cancer  -
XX
PS   Claim 4; Page 182; 193pp; English.
XX
CC   The present invention describes polypeptides (I) comprising an
CC   immunogenic portion of a native Wilm's Tumour gene product polypeptide,
CC   WT1, (or variants of the immunogenic portion retaining the ability to
CC   react with WT1-specific antisera and/or T-cell lines or clones) and
CC   comprising 16 consecutive amino acids (aa) or less of a native WT1
CC   polypeptide. The polypeptides are useful therapeutically and to
CC   manufacture medicaments for enhancing/inducing an immune response in
CC   patients. The polypeptides, mimetics or polynucleotides can be included
CC   with a carrier/excipient in pharmaceutical compositions or with a
CC   non-specific immune response enhancer (e.g. an adjuvant or enhancer
CC   preferentially enhancing a T cell response) in vaccines. Pharmaceutical
CC   compositions and vaccines can be administered to human patients to
CC   enhance or induce an immune response specific for WT1 or a cell
CC   expressing WT1, useful to inhibit the development of malignant diseases
CC   associated with WT1 expression, e.g. leukemia (especially acute/chronic
CC   myeloid leukemia or acute lymphocytic leukemia) or cancer (especially
CC   breast, lung, thyroid or gastrointestinal cancer, or a melanoma). Y98501
CC   to Y98811 represent polypeptide sequences, and A13848 to A13862 represent
CC   PCR primers, used in the exemplification of the present invention.
XX
SQ   Sequence   9 AA;


  Query Match             47.8%;  Score 22;  DB 21;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 NVQDLN 8
           :|:|||
Db       1 dvrdln 6


RESULT  25
Y98783
ID   Y98783 standard; Peptide; 9 AA.
XX
AC   Y98783;
XX
DT   31-JUL-2000  (first entry)
XX
DE   WT1 derived immunogenic peptide SEQ ID NO:298.
XX
KW   WT1; immunotherapy; immunogenic; malignant disease; cancer; leukaemia;
KW   metastatic disease; mouse; human; Wilm's tumour; immune response;
KW   vaccine.
XX
OS   Mus musculus.
XX
PN   WO200018795-A2.
XX
PD   06-APR-2000.
XX
PF   30-SEP-1999;   99WO-US22819.
XX
PR   30-SEP-1998;   98US-0164223.
PR   25-MAR-1999;   99US-0276484.
```

```
XX
PA   (CORI-) CORIXA CORP.
PA   (GAIG/) GAIGER A.
XX
PI   Gaiger A,  Cheever M;
XX
DR   WPI; 2000-293107/25.
XX
PT   Novel polypeptides comprising an immunogenic portion of a native WT1
PT   polypeptide, useful for inhibiting the development of malignant
PT   diseases associated with WT1 expression e.g. leukemia or cancer  -
XX
PS   Claim 4; Page 187; 193pp; English.
XX
CC   The present invention describes polypeptides (I) comprising an
CC   immunogenic portion of a native Wilm's Tumour gene product polypeptide,
CC   WT1, (or variants of the immunogenic portion retaining the ability to
CC   react with WT1-specific antisera and/or T-cell lines or clones) and
CC   comprising 16 consecutive amino acids (aa) or less of a native WT1
CC   polypeptide. The polypeptides are useful therapeutically and to
CC   manufacture medicaments for enhancing/inducing an immune response in
CC   patients. The polypeptides, mimetics or polynucleotides can be included
CC   with a carrier/excipient in pharmaceutical compositions or with a
CC   non-specific immune response enhancer (e.g. an adjuvant or enhancer
CC   preferentially enhancing a T cell response) in vaccines. Pharmaceutical
CC   compositions and vaccines can be administered to human patients to
CC   enhance or induce an immune response specific for WT1 or a cell
CC   expressing WT1, useful to inhibit the development of malignant diseases
CC   associated with WT1 expression, e.g. leukemia (especially acute/chronic
CC   myeloid leukemia or acute lymphocytic leukemia) or cancer (especially
CC   breast, lung, thyroid or gastrointestinal cancer, or a melanoma). Y98501
CC   to Y98811 represent polypeptide sequences, and A13848 to A13862 represent
CC   PCR primers, used in the exemplification of the present invention.
XX
SQ   Sequence   9 AA;


  Query Match             47.8%;  Score 22;  DB 21;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 NVQDLN 8
              :|:|||
Db          2 dvrdln 7


RESULT  26
W61747
ID   W61747 standard; peptide; 6 AA.
XX
AC   W61747;
XX
DT   09-NOV-1998  (first entry)
XX
DE   Peptide mimetic of cytokine receptor gamma chain 14.
XX
KW   Peptide mimetic; gamma-chain; cytokine receptor; signal transduction;
KW   autoimmune disease; graft vs. host disease; transplant rejection;
KW   graft rejection; interleukin; immunosuppressant; T cell; B cell.
XX
OS   Synthetic.
XX
PN   WO9834631-A1.
XX
PD   13-AUG-1998.
XX
PF   06-FEB-1998;   98WO-US02339.
XX
PR   07-FEB-1997;   97US-0036941.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Korngold R,  Townsend RM;
```

```
XX
DR   WPI; 1998-446944/38.
XX
PT   New peptide(s) mimicking a loop in the gamma chain of cytokine
PT   receptors - inhibit signal transduction through these receptors,
PT   useful as immunosuppressants for treating or preventing e.g.
PT   leukaemia, autoimmune disease, graft rejection
XX
PS   Claim 2; Page 28; 54pp; English.
XX
CC   The peptide mimetics W61734-W61756 and W77001-W77014 mimic a loop on the
CC   gamma-chain of cytokine receptors, and interacts with a cytokine or a
CC   gamma-chain partner receptor chain of a heterodimeric cytokine receptor.
CC   They inhibit signal transduction mediated by cytokine:receptor binding
CC   (of cytokines that bind to receptors with a gamma-chain).  They are used
CC   to inhibit or suppress cytokine-mediated immune responses, growth,
CC   proliferation, function and activity of cells. Particularly they are used
CC   for treatment or prevention of lymphoma, leukaemia, allergy
CC   (e.g. asthma), autoimmune disease (e.g. rheumatoid arthritis, lupus,
CC   multiple sclerosis or myasthenia gravis), graft vs. host disease and
CC   transplant or graft rejection.  They inhibit function of interleukin
CC   (IL)-2, -4, -7, -9, -13 and -15 (the receptors for which include a common
CC   gamma -chain), so function as immunosuppressants by reducing
CC   proliferation of T and B cells.
XX
SQ   Sequence   6 AA;


  Query Match             45.7%;  Score 21;  DB 19;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LNVQDL 7
              ||:|:|
Db          1 lnlqnl 6


RESULT  27
Y63740
ID   Y63740 standard; Peptide; 7 AA.
XX
AC   Y63740;
XX
DT   02-MAR-2000  (first entry)
XX
DE   Desmoglein cell adhesion recognition cyclic peptide SEQ ID NO:3192.
XX
KW   Modulation; nonclassical cadherin mediated cell adhesion; CAR;
KW   inhibition; cadherin extracellular domain; cell adhesion recognition;
KW   OB-cadherin; cadherin-5; cadherin-6; cadherin-7; cadherin-8;
KW   cadherin-12; cadherin-14; cadherin-15; T-cadherin; PB-cadherin;
KW   cadherin related neuronal receptor; LI-cadherin; protocadherin;
KW   desmoglein; desmocollin; calcium binding; cancer; tumour; obesity;
KW   rheumatoid arthritis; multiple sclerosis; diabetes; metastasis;
KW   neurological disease; cyclic.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1..7
FT                   /note= "the terminal residues are condensed with each
FT                          other to form a cyclic peptide"
XX
PN   WO9957149-A2.
XX
PD   11-NOV-1999.
XX
PF   05-MAY-1999;   99WO-CA00363.
XX
PR   05-MAY-1998;   98US-0073040.
PR   06-NOV-1998;   98US-0187859.
PR   20-JAN-1999;   99US-0234395.
```

```
PR   08-MAR-1999;   99US-0264516.
XX
PA   (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI   Blaschuk OW,  Gour BJ,  Byers S;
XX
DR   WPI; 2000-038791/03.
XX
PT   New cadherin modulating agents, used for modulating nonclassical
PT   cadherin-mediated functions for treating e.g. cancers, obesity,
PT   rheumatoid arthritis, multiple sclerosis, diabetes or a neurological
PT   disease  -
XX
PS   Claim 90; Page 208; 252pp; English.
XX
CC   The present invention describes cadherin modulating agents (MA)
CC   comprising peptides which comprise a nonclassical cadherin cell adhesion
CC   recognition (CAR) sequence. The MAs can be used for modulating
CC   nonclassical cadherin-mediated functions. They can be used for e.g.
CC   inhibiting adhesion of nonclassical-cadherin expressing cells in a
CC   mammal, enhancing delivery of a drug through the skin of a mammal,
CC   enhancing delivery of a drug to a tumour in a mammal, treating cancer in
CC   a mammal, inhibiting metastasis of a cancer in a mammal, inhibiting
CC   angiogenesis in a mammal, inducing apoptosis in a nonclassical cadherin-
CC   expressing cell, preventing or treating obesity in a mammal, stimulating
CC   blood vessel regression in a mammal, enhancing drug delivery to the
CC   central nervous system, treating a demyelinating neurological disease,
CC   increasing vasopermeability in a mammal, enhancing adhesion of
CC   nonclassical cadherin-expressing cells, inhibiting synaptic stability in
CC   a mammal, or preventing pregnancy in a mammal. They can also be used for
CC   e.g. enhancing or directing neurite outgrowth, facilitating wound healing
CC   or reducing scar tissue, or enhancing adhesion of foreign tissue in a
CC   mammal. They can also be used for treating e.g. psoriasis, arthritis,
CC   age-related macular degeneration, multiple sclerosis and diabetes. The
CC   products can also be used for detection and diagnosis and in
CC   bioreactors. Y60592 to Y64572 represent specifically claimed peptides,
CC   and Y64573 to Y64643 and Z33183 to Z33186 represent sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence   7 AA;

  Query Match             45.7%;  Score 21;  DB 21;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLNVQD 6
            |||  |
Db        2 ylnkdd 7


RESULT  28
W77008
ID   W77008 standard; peptide; 8 AA.
XX
AC   W77008;
XX
DT   09-NOV-1998  (first entry)
XX
DE   Peptide mimetic of cytokine receptor gamma chain 31.
XX
KW   Peptide mimetic; gamma-chain; cytokine receptor; signal transduction;
KW   autoimmune disease; graft vs. host disease; transplant rejection;
KW   graft rejection; interleukin; immunosuppressant; T cell; B cell.
XX
OS   Synthetic.
XX
PN   WO9834631-A1.
XX
PD   13-AUG-1998.
XX
PF   06-FEB-1998;   98WO-US02339.
XX
```

```
PR   07-FEB-1997;   97US-0036941.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Korngold R,  Townsend RM;
XX
DR   WPI; 1998-446944/38.
XX
PT   New peptide(s) mimicking a loop in the gamma chain of cytokine
PT   receptors - inhibit signal transduction through these receptors,
PT   useful as immunosuppressants for treating or preventing e.g.
PT   leukaemia, autoimmune disease, graft rejection
XX
PS   Claim 8; Page 29; 54pp; English.
XX
CC   The peptide mimetics W61734-W61756 and W77001-W77014 mimic a loop on the
CC   gamma-chain of cytokine receptors, and interacts with a cytokine or a
CC   gamma-chain partner receptor chain of a heterodimeric cytokine receptor.
CC   They inhibit signal transduction mediated by cytokine:receptor binding
CC   (of cytokines that bind to receptors with a gamma-chain).  They are used
CC   to inhibit or suppress cytokine-mediated immune responses, growth,
CC   proliferation, function and activity of cells. Particularly they are used
CC   for treatment or prevention of lymphoma, leukaemia, allergy
CC   (e.g. asthma), autoimmune disease (e.g. rheumatoid arthritis, lupus,
CC   multiple sclerosis or myasthenia gravis), graft vs. host disease and
CC   transplant or graft rejection.  They inhibit function of interleukin
CC   (IL)-2, -4, -7, -9, -13 and -15 (the receptors for which include a common
CC   gamma -chain), so function as immunosuppressants by reducing
CC   proliferation of T and B cells.
XX
SQ   Sequence   8 AA;

  Query Match             45.7%;  Score 21;  DB 19;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LNVQDL 7
            ||:|:|
Db        2 lnlqnl 7


RESULT  29
Y63743
ID   Y63743 standard; Peptide; 8 AA.
XX
AC   Y63743;
XX
DT   02-MAR-2000  (first entry)
XX
DE   Desmoglein cell adhesion recognition cyclic peptide SEQ ID NO:3195.
XX
KW   Modulation; nonclassical cadherin mediated cell adhesion; CAR;
KW   inhibition; cadherin extracellular domain; cell adhesion recognition;
KW   OB-cadherin; cadherin-5; cadherin-6; cadherin-7; cadherin-8;
KW   cadherin-12; cadherin-14; cadherin-15; T-cadherin; PB-cadherin;
KW   cadherin related neuronal receptor; LI-cadherin; protocadherin;
KW   desmoglein; desmocollin; calcium binding; cancer; tumour; obesity;
KW   rheumatoid arthritis; multiple sclerosis; diabetes; metastasis;
KW   neurological disease; cyclic.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1..8
FT                   /note= "the terminal residues are condensed with each
FT                          other to form a cyclic peptide"
XX
PN   WO9957149-A2.
XX
PD   11-NOV-1999.
XX
```

```
PF   05-MAY-1999;   99WO-CA00363.
XX
PR   05-MAY-1998;   98US-0073040.
PR   06-NOV-1998;   98US-0187859.
PR   20-JAN-1999;   99US-0234395.
PR   08-MAR-1999;   99US-0264516.
XX
PA   (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI   Blaschuk OW,  Gour BJ,  Byers S;
XX
DR   WPI; 2000-038791/03.
XX
PT   New cadherin modulating agents, used for modulating nonclassical
PT   cadherin-mediated functions for treating e.g. cancers, obesity,
PT   rheumatoid arthritis, multiple sclerosis, diabetes or a neurological
PT   disease  -
XX
PS   Claim 90; Page 208; 252pp; English.
XX
CC   The present invention describes cadherin modulating agents (MA)
CC   comprising peptides which comprise a nonclassical cadherin cell adhesion
CC   recognition (CAR) sequence. The MAs can be used for modulating
CC   nonclassical cadherin-mediated functions. They can be used for e.g.
CC   inhibiting adhesion of nonclassical-cadherin expressing cells in a
CC   mammal, enhancing delivery of a drug through the skin of a mammal,
CC   enhancing delivery of a drug to a tumour in a mammal, treating cancer in
CC   a mammal, inhibiting metastasis of a cancer in a mammal, inhibiting
CC   angiogenesis in a mammal, inducing apoptosis in a nonclassical cadherin-
CC   expressing cell, preventing or treating obesity in a mammal, stimulating
CC   blood vessel regression in a mammal, enhancing drug delivery to the
CC   central nervous system, treating a demyelinating neurological disease,
CC   increasing vasopermeability in a mammal, enhancing adhesion of
CC   nonclassical cadherin-expressing cells, inhibiting synaptic stability in
CC   a mammal, or preventing pregnancy in a mammal. They can also be used for
CC   e.g. enhancing or directing neurite outgrowth, facilitating wound healing
CC   or reducing scar tissue, or enhancing adhesion of foreign tissue in a
CC   mammal. They can also be used for treating e.g. psoriasis, arthritis,
CC   age-related macular degeneration, multiple sclerosis and diabetes. The
CC   products can also be used for detection and diagnosis and in
CC   bioreactors. Y60592 to Y64572 represent specifically claimed peptides,
CC   and Y64573 to Y64643 and Z33183 to Z33186 represent sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence   8 AA;


  Query Match             45.7%;  Score 21;  DB 21;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLNVQD 6
           |||  |
Db       3 ylnkdd 8


RESULT  30
R59112
ID   R59112 standard; peptide; 9 AA.
XX
AC   R59112;
XX
DT   02-MAY-1995  (first entry)
XX
DE   Peptide fragment (1.0384) of HBV binds HLA-A2.1.
XX
KW   antigen; epitope; immunogenic target protein; PSA; HBVc; HBVs; EBV;
KW   HIV1; core antigen; surface antigen; pharmaceutical composition;
KW   in vivo; ex vivo; therapeutic; diagnostic; MHC class I molecule;
KW   major histocompatability complex; HLA-A2.1; 9mer; 10mer; anchor;
KW   human leukocyte antigen.
XX
OS   Hepatitis B virus strain adw.
```

```
XX
PN   WO9420127-A.
XX
PD   15-SEP-1994.
XX
PF   04-MAR-1994;   94WO-US02353.
XX
PR   05-MAR-1993;   93US-0027146.
PR   04-JUN-1993;   93US-0073205.
PR   29-NOV-1993;   93US-0159184.
XX
PA   (CYTE-) CYTEL CORP.
XX
PI   Grey HM,  Kast WM,  Sette A,  Sidney J;
XX
DR   WPI; 1994-302678/37.
XX
PT   Immunogenic peptide(s) having an HLA-A2.1 binding motif - used
PT   for treatment or prophylaxis of cancer, virus infection or
PT   autoimmune diseases.
XX
PS   Example 5; Page 100; 138pp; English.
XX
CC   R59104-264 are immunogenic 9mer peptides that contain a HLA-A2.1 binding
CC   motif. These peptides bind HLA-A2.1 and have a binding affinity of at
CC   least 1% as compared to a reference peptide (R71293). R59112 has an IC50
CC   of 0.29 and the sequence occurs at position 1279 in the HBV POL protein.
CC   The peptides of the invention can induce cytotoxic T lymphocytes which
CC   can react with target cells. They can be used for the treatment or
CC   prophlyaxis of cancer, eg. prostate cancer or lymphoma, etc.
XX
SQ   Sequence   9 AA;


  Query Match             45.7%;  Score 21;  DB 15;  Length 9;
  Best Local Similarity   55.6%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 YLNVQDLNL 9
           :|  | |||
Db       1 fltkqylnl 9


RESULT  31
R59233
ID   R59233 standard; peptide; 9 AA.
XX
AC   R59233;
XX
DT   04-MAY-1995  (first entry)
XX
DE   Peptide fragment (1.0164) of HIV binds HLA-A2.1.
XX
KW   antigen; epitope; immunogenic target protein; PSA; HBVc; HBVs; EBV;
KW   HIV1; core antigen; surface antigen; pharmaceutical composition;
KW   in vivo; ex vivo; therapeutic; diagnostic; MHC class I molecule;
KW   major histocompatability complex; HLA-A2.1; 9mer; 10mer; anchor;
KW   human leukocyte antigen.
XX
OS   Human immunodeficiency virus.
XX
PN   WO9420127-A.
XX
PD   15-SEP-1994.
XX
PF   04-MAR-1994;   94WO-US02353.
XX
PR   05-MAR-1993;   93US-0027146.
PR   04-JUN-1993;   93US-0073205.
PR   29-NOV-1993;   93US-0159184.
XX
PA   (CYTE-) CYTEL CORP.
XX
```

```
PI   Grey HM,  Kast WM,  Sette A,  Sidney J;
XX
DR   WPI; 1994-302678/37.
XX
PT   Immunogenic peptide(s) having an HLA-A2.1 binding motif - used
PT   for treatment or prophylaxis of cancer, virus infection or
PT   autoimmune diseases.
XX
PS   Example 5; Page 105; 138pp; English.
XX
CC   R59104-264 are immunogenic 9mer peptides that contain a HLA-A2.1 binding
CC   motif. These peptides bind HLA-A2.1 and have a binding affinity of at
CC   least 1% as compared to a reference peptide (R71293). R59233 has an IC50
CC   of 0 and the sequence occurs at position 614 in the HIV POL protein.
CC   The peptides of the invention can induce cytotoxic T lymphocytes which
CC   can react with target cells. They can be used for the treatment or
CC   prophlyaxis of cancer, eg. prostate cancer or lymphoma, etc.
XX
SQ   Sequence   9 AA;


  Query Match             45.7%;  Score 21;  DB 15;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        2 LNVQDLNL 9
            ||: :||:
Db        2 lnlgnlnv 9


RESULT  32
R70596
ID   R70596 standard; Peptide; 9 AA.
XX
AC   R70596;
XX
DT   14-FEB-1996  (first entry)
XX
DE   HIV(B35)GAG-8, human immunodeficiency virus epitope.
XX
KW   HLA; human lymphocyte antigen; HIV; human immunodeficiency virus;
KW   binding peptide; induce killer cell; prevention; treatment; AIDS;
KW   autoimmune disease syndrome; vaccine.
XX
OS   Human immunodeficiency virus.
XX
PN   WO9511255-A1.
XX
PD   27-APR-1995.
XX
PF   19-OCT-1994;   94WO-JP01756.
XX
PR   19-OCT-1993;   93JP-0261302.
XX
PA   (AJIN ) AJINOMOTO CO INC.
PA   (AJIN ) AJINOMOTO KK.
XX
PI   Miwa K,  Takiguchi M;
XX
DR   WPI; 1995-170188/22.
XX
PT   HLA-binding peptide fragments from HIV proteins - induce killer
PT   cells which target HIV-infected cells and can be incorporated into
PT   anti-HIV vaccines
XX
PS   Example 1; Page 10; 61pp; Japanese.
XX
CC   R70596 is a peptide fragment derived from an HIV (Human Immunodeficiency
CC   Virus) protein and is capable of binding to a human lymphocyte antigen.
CC   The peptide can induce killer cells which target HIV-infected cells.
CC   It is also useful in the prevention and treatment of HIV and AIDS.
CC   Anti-HIV vaccines may incorporate the peptides, or may incorporate a
CC   vector (such as vaccinia or BCG) contg. DNA encoding the peptides.
```

```
XX
SQ   Sequence   9 AA;


  Query Match             45.7%;  Score 21;  DB 16;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 QDLN 8
            ||||
Db        3 qdln 6


RESULT  33
R70067
ID   R70067 standard; peptide; 9 AA.
XX
AC   R70067;
XX
DT   06-OCT-1995  (first entry)
XX
DE   Control peptide 927.30 for HBV epitope stimulation of T-cells.
XX
KW   Hepatitis B virus polymerase; cytotoxic T cell response; prophylactic;
KW   vaccine; chronic; acute HBV infection; carrier.
XX
OS   Hepatitis B virus.
XX
PN   WO9503777-A.
XX
PD   09-FEB-1995.
XX
PF   01-AUG-1994;   94WO-US08685.
XX
PR   02-AUG-1993;   93US-0100870.
XX
PA   (SCRI ) SCRIPPS RES INST.
XX
PI   Chisari FV;
XX
DR   WPI; 1995-082004/11.
XX
PT   New peptides inducing cytotoxic T lymphocytes to hepatitis B
PT   virus  - are regions of HB polymerase protein, for treating acute
PT   and chronic infections
XX
PS   Example 1; Page 35; 85pp; English.
XX
CC   Peptides R70066-7 are control peptides in an assay for stimulation of
CC   cytotoxic T cells (CTLs) by epitopes from the hepatitis B virus (HBV)
CC   polymerase (HBpol) (see R70044-59).  The HBpol peptides can be used,
CC   prophylactically as vaccines, together with, or conjugated to, HBV helper
CC   epitopes (R70060-4).  The peptides can be used, particularly ex vivo, to
CC   stimulate CTL cells,which cells can be reintroduced into patients who
CC   have chronic or acute HBV infections or are carriers, especially in
CC   treatments to prevent conversion from acute to chronic infections.
XX
SQ   Sequence   9 AA;


  Query Match             45.7%;  Score 21;  DB 16;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        2 LNVQDLNL 9
            ||: :||:
Db        2 lnlgnlnv 9


RESULT  34
Y07297
ID   Y07297 standard; peptide; 9 AA.
XX
```

```
AC   Y07297;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Peptide against SH-PTP2 SH2 domain.
XX
KW   O-Malonyl-tyrosine; derivative; signal transduction; binding; Src;
KW   P1-3 kinase; SH2 domain; Grb2; SH-PTP2; ligand; cancer; therapy;
KW   diabetes; immune dysfunction.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "Acylated N-terminus"
FT   Modified-site   2
FT                   /note= "O-malonyltyrosyl derivative"
FT   Modified-site   6
FT                   /note= "amidated C-terminus"
XX
PN   WO9630332-A1.
XX
PD   03-OCT-1996.
XX
PF   29-MAR-1996;   96WO-US04311.
XX
PR   31-MAR-1995;   95US-0414520.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Akamatsu M,  Burke TR,  Kole HK,  Roller PR,  Yan X;
PI   Ye B;
XX
DR   WPI; 1996-455230/45.
XX
PT   New O-malonyl-tyrosine derivs. - useful as protein-tyrosine
PT   phosphatase inhibitors
XX
PS   Claim 21; Page 55; 66pp; English.
XX
CC   The invention relates to the generation of peptides containing
CC   O-Malonyl-tyrosine derivatives (Y0294-Y09296). The peptides are useful
CC   for disease treatment or prevention by affecting signal transduction
CC   pathways of cells. e.g. to inhibit binding interactions of the P1-3
CC   kinase C-terminal p85 SH2 domain, the Src SH2 domain, the Grb2 SH2
CC   domain or the N-terminal SH-PTP2 SH2 domain with ligands and are useful
CC   for treatment or prevention of diseases, especially cancer (optionally
CC   in combination with chemo- or radiotherapy), diabetes or immune
CC   dysfunction.
XX
SQ   Sequence   9 AA;

  Query Match             45.7%;  Score 21;  DB 17;  Length 9;
  Best Local Similarity   62.5%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 LNVQDLNL 9
              ||  ||:|
Db          1 lnxidldl 8


RESULT  35
W78574
ID   W78574 standard; peptide; 9 AA.
XX
AC   W78574;
XX
DT   04-NOV-1998  (first entry)
XX
DE   SH2 domain binding inhibiting peptide SEQ ID NO:60.
XX
KW   SH2 domain; binding; inhibition; interaction; site specific;
```

```
KW   signal transduction; protein tyrosine kinase; phosphotyrosine;
KW   growth factor receptor; oncogene; cellular growth; cell proliferation;
KW   metabolic control; diabetes; PTK; proto-oncogene; insulin receptor.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 3
FT                   /note= "unspecified"
XX
PN   US5801149-A.
XX
PD   01-SEP-1998.
XX
PF   21-MAR-1995;   95US-0408604.
XX
PR   21-MAR-1995;   95US-0408604.
PR   19-JUN-1991;   91US-0722359.
PR   09-OCT-1992;   92US-0959949.
PR   08-OCT-1993;   93US-0134558.
XX
PA   (JOSL-) JOSLIN DIABETES CENT INC.
XX
PI   Shoelson S;
XX
DR   WPI; 1998-494822/42.
XX
PT   Inhibiting site-specific SH2 domain interaction - with peptide
PT   containing phosphotyrosine or phosphotyrosine mimic
XX
PS   Disclosure; Column 53; 70pp; English.
XX
CC   A method has been developed of inhibiting a site-specific interaction
CC   between a first molecule having an SH2 domain and a second molecule that
CC   interacts with the SH2 domain. The method comprises contacting the first
CC   molecule with a 4- to 30-mer peptide containing a sequence of formula:
CC   R1-Met-R3-Met  (I), where R1 = phosphotyrosine (pTyr) or a
CC   phosphotyrosine analogue having a hydrolysis-resistant phosphorous
CC   moiety, and R3 = any amino acid. W78501 to W78523 represent specifically
CC   claimed examples of the peptides described. The peptides are useful for
CC   modulating both cellular growth to control unwanted cell proliferation
CC   in e.g selected malignancies and for metabolic control in e.g. diabetes,
CC   by inhibiting signal transduction molecules such as protein tyrosine
CC   kinases (PTKs) which include growth factor receptors, proto-oncogene and
CC   oncogene products and the insulin receptor. The peptides are also useful
CC   for treating and for studying the enzymatic mechanisms of PTPase
CC   activity and the metabolic and biochemical roles of PTPases. W78524 to
CC   W78702 represent other peptides given in the present invention, but
CC   which are not specifically claimed.
XX
SQ   Sequence   9 AA;

  Query Match             45.7%;  Score 21;  DB 19;  Length 9;
  Best Local Similarity   62.5%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 LNVQDLNL 9
              ||  ||:|
Db          1 lnxidldl 8


Search completed: December 16, 2000, 03:07:20
Job time: 8061 sec
```

```
                     GenCore version 4.5
            Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 02:00:58 ; Search time 113.2 Seconds
                                              (without alignments)
                                              7.424 Million cell updates/sec

Title:          US-09-529-121-5
Perfect score:  48
Sequence:       1 YLSGACLNL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters:      297973

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      SPTREMBL_14:*
               1:  sp_archea:*
               2:  sp_bacteria:*
               3:  sp_fungi:*
               4:  sp_human:*
               5:  sp_invertebrate:*
               6:  sp_mammal:*
               7:  sp_mhc:*
               8:  sp_organelle:*
               9:  sp_phage:*
               10:  sp_plant:*
               11:  sp_rodent:*
               12:  sp_virus:*
               13:  sp_vertebrate:*
               14:  sp_unclassified:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                %
Result         Query
  No.   Score  Match Length DB  ID                 Description
 ------------------------------------------------------------------------------
    1    36.5   76.0    344  5  Q9VED2             Q9ved2 drosophila
    2    36     75.0   1624  5  Q9V3K8             Q9v3k8 drosophila
    3    35     72.9   3638  4  Q15142             Q15142 homo sapien
    4    35     72.9   4292  4  Q15141             Q15141 homo sapien
    5    35     72.9   4302  4  Q15140             Q15140 homo sapien
    6    34     70.8     91  7  Q31147             Q31147 mus musculu
    7    34     70.8    265 10  Q9SHT3              Q9sht3 arabidopsis
    8    34     70.8    375  4  O75098             O75098 homo sapien
    9    34     70.8   2218 12  O09705              O09705 lassa virus
   10    34     70.8   2395  5  Q27167             Q27167 paramecium
   11    33     68.8     85 12  O37322              O37322 heliothis a
   12    33     68.8    114  5  Q22960             Q22960 caenorhabdi
   13    33     68.8    348  3  O94038             O94038 candida alb
   14    33     68.8    362  7  Q9TP68             Q9tp68 homo sapien
   15    33     68.8    363  3  O74484             O74484 schizosacch
   16    33     68.8    453 10  P93818              P93818 arabidopsis
   17    33     68.8    514  2  Q9XB13             Q9xb13 mycobacteri
   18    33     68.8    648 10  Q41716              Q41716 xanthium sp
   19    33     68.8    648 10  Q41717              Q41717 xanthium sp
```

```
   20    33     68.8    659 10  Q42767             Q42767 gossypium h
   21    33     68.8    659 10  Q42768             Q42768 gossypium h
   22    33     68.8    666 10  Q9ZSU3             Q9zsu3 bassia scop
   23    33     68.8    933  2  O25380             O25380 helicobacte
   24    33     68.8   1017  3  O74670             O74670 pneumocysti
   25    33     68.8   2209 12  Q9YZ25              Q9yz25 lymphocytic
   26    32     66.7    136  7  Q9TPZ1             Q9tpz1 ambystoma m
   27    32     66.7    146 12  Q89173              Q89173 variola vir
   28    32     66.7    168  8  Q36114             Q36114 trichoniscu
   29    32     66.7    221  8  Q36113             Q36113 trichoniscu
   30    32     66.7    250  5  Q9U8V4             Q9u8v4 eptatretus
   31    32     66.7    250 11  Q9WVJ5              Q9wvj5 mus musculu
   32    32     66.7    286 10  Q9S9P2              Q9s9p2 arabidopsis
   33    32     66.7    300  2  O07868             O07868 streptococc
   34    32     66.7    303  4  O95327             O95327 homo sapien
   35    32     66.7    323 11  O09005              O09005 mus musculu
   36    32     66.7    334  7  Q30770             Q30770 ovis aries
   37    32     66.7    362  3  O93827             O93827 candida alb
   38    32     66.7    364  3  O74624             O74624 trichoderma
   39    32     66.7    370 10  Q9SUR1              Q9sur1 arabidopsis
   40    32     66.7    378  5  O61715             O61715 caenorhabdi
   41    32     66.7    404  5  O76258             O76258 caenorhabdi
   42    32     66.7    538 10  O23174              O23174 arabidopsis
   43    32     66.7    589 10  Q9ZQF7              Q9zqf7 arabidopsis
   44    32     66.7    604  2  Q9RMC9             Q9rmc9 acinetobact
   45    32     66.7    620  1  O93676             O93676 sulfolobus

                              ALIGNMENTS

RESULT   1
Q9VED2
ID   Q9VED2      PRELIMINARY;      PRT;   344 AA.
AC   Q9VED2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG14316 PROTEIN.
GN   CG14316.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE; 20196006.
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
```

```
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003720; AAF55493.1; -.
DR   FLYBASE; FBgn0038567; CG14316.
SQ   SEQUENCE   344 AA;  39318 MW;  3ECAA947157719C7 CRC64;

  Query Match             76.0%;  Score 36.5;  DB 5;  Length 344;
  Best Local Similarity   75.0%;  Pred. No. 12;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    1;

Qy        1 YLSGA---CLNL 9
              |||||   ||||
Db      134 YLSGAPLKCLNL 145

RESULT   2
Q9V3K8
ID   Q9V3K8      PRELIMINARY;      PRT;  1624 AA.
AC   Q9V3K8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CG2995 PROTEIN.
GN   CG2995 OR EG:BACR37P7.2.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE; 20196006.
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
```

```
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Madueno E., de Pablos B., Modolell J.;
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003417; AAF45487.1; -.
DR   EMBL; AL050231; CAB65850.1; -.
DR   FLYBASE; FBgn0029516; CG2995.
DR   INTERPRO; IPR001214; -.
DR   INTERPRO; IPR002110; -.
DR   PFAM; PF00023; ank; 5.
DR   PFAM; PF00856; SET; 1.
SQ   SEQUENCE   1624 AA;  179964 MW;  AAA34BAFCDAE5121 CRC64;

  Query Match             75.0%;  Score 36;  DB 5;  Length 1624;
  Best Local Similarity   66.7%;  Pred. No. 68;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGACLNL 9
              ||:| ||:|
Db     1084 YLNGTCLHL 1092

RESULT   3
Q15142
ID   Q15142      PRELIMINARY;      PRT;  3638 AA.
AC   Q15142;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   POLYCYSTIC KIDNEY DISEASE-ASSOCIATED PROTEIN.
GN   PKD1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94273192.
RA   Ward C.J., Peral B., Hughes J., Thomas S., Gamble V., MacCarthy A.B.,
RA   Sloane-Stanley J., Buckle V.J., Kearney L., Higgs D.R.,
RA   Ratcliffe P.J., Harris P.C., Roelfsema J.H., Spruit L.L., Saris J.J.,
RA   Dauwerse H.G., Peters D.J.M., Breuning M.H., Nellist M.,
RA   Brook-Carter P.T., Maheshwar M.M., Cordeiro I., Santos H., Cabral P.,
RA   Sampson J.R., Janssen B., Hesseling-Janssen A.L.W.,
RA   den Ouweland A.M.W., Eussen B., Verhoef S., Lindhout D.,
RA   Halley D.J.J.;
RT   "The polycystic kidney disease 1 gene encodes a 14 kb transcript and
RT   lies within a duplicated region on chromosome 16. The European
RT   Polycystic Kidney Disease Consortium [published errata appear in Cell
RT   1994 Aug 26;78(4):following 724 and 1995 Jun 30;81(7):following
RT   1170].";
RL   Cell 77:881-894(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95359960.
RA   Burn T.C., Connors T.D., Dackowski W.R., Petry L.R., Van Raay T.J.,
RA   Millholland J., Venet M., Miller G., Hakim R.M., Doggett N.A.,
RA   Landes G.M., Klinger K.W., Qian F., Onuchic L.F., Watnick T.,
RA   Germino G.G.;
```

```
RT   "Analysis of the genomic sequence for the autosomal dominant
RT   polycystic kidney disease (PKD1) gene predicts the presence of a
RT   leucine-rich repeat. The American PKD1 Consortium (APKD1
RT   Consortium).";
RL   Hum. Mol. Genet. 4:575-582(1995).
DR   EMBL; L39891; AAB59488.1; -.
DR   INTERPRO; IPR000203; -.
DR   INTERPRO; IPR000372; -.
DR   INTERPRO; IPR000434; -.
DR   INTERPRO; IPR000483; -.
DR   INTERPRO; IPR000601; -.
DR   INTERPRO; IPR001024; -.
DR   INTERPRO; IPR001304; -.
DR   INTERPRO; IPR001611; -.
DR   INTERPRO; IPR002859; -.
DR   PFAM; PF00059; lectin_c; 1.
DR   PFAM; PF00560; LRR; 1.
DR   PFAM; PF00801; PKD; 11.
DR   PFAM; PF01462; LRRNT; 1.
DR   PFAM; PF01463; LRRCT; 1.
DR   PFAM; PF01477; PLAT; 1.
DR   PFAM; PF01825; GPS; 1.
DR   PFAM; PF02010; REJ; 1.
DR   PRINTS; PR00500; POLYCYSTIN.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
SQ   SEQUENCE   3638 AA;  392182 MW;  8267F06B5E04B0D1 CRC64;

  Query Match             72.9%;  Score 35;  DB 4;  Length 3638;
  Best Local Similarity   75.0%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGACLN 8
           || | |||
Db    1740 YLEGRCLN 1747


RESULT   4
Q15141
ID   Q15141      PRELIMINARY;      PRT;  4292 AA.
AC   Q15141;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   POLYCYSTIC KIDNEY DISEASE 1 PROTEIN.
GN   PKD1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94273192.
RA   Ward C.J., Peral B., Hughes J., Thomas S., Gamble V., MacCarthy A.B.,
RA   Sloane-Stanley J., Buckle V.J., Kearney L., Higgs D.R.,
RA   Ratcliffe P.J., Harris P.C., Roelfsema J.H., Spruit L.L., Saris J.J.,
RA   Dauwerse H.G., Peters D.J.M., Breuning M.H., Nellist M.,
RA   Brook-Carter P.T., Maheshwar M.M., Cordeiro I., Santos H., Cabral P.,
RA   Sampson J.R., Janssen B., Hesseling-Janssen A.L.W.,
RA   den Ouweland A.M.W., Eussen B., Verhoef S., Lindhout D.,
RA   Halley D.J.J.;
RT   "The polycystic kidney disease 1 gene encodes a 14 kb transcript and
RT   lies within a duplicated region on chromosome 16. The European
RT   Polycystic Kidney Disease Consortium [published errata appear in Cell
RT   1994 Aug 26;78(4):following 724 and 1995 Jun 30;81(7):following
RT   1170].";
RL   Cell 77:881-894(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95392566.
RA   Hughes J., Ward C.J., Peral B., Aspinwall R., Clark K., Millan J.L.,
RA   Gamble V., Harris P.C.;
RT   "The polycystic kidney disease 1 (PKD1) gene encodes a novel protein
RT   with multiple cell recognition domains.";
```

```
RL   Nat. Genet. 10:151-160(1995).
DR   EMBL; L43619; AAC41765.1; -.
DR   EMBL; L43601; AAC41765.1; JOINED.
DR   EMBL; L43602; AAC41765.1; JOINED.
DR   EMBL; L43604; AAC41765.1; JOINED.
DR   EMBL; L43605; AAC41765.1; JOINED.
DR   EMBL; L43610; AAC41765.1; JOINED.
DR   EMBL; L43617; AAC41765.1; JOINED.
DR   EMBL; L43618; AAC41765.1; JOINED.
DR   INTERPRO; IPR000203; -.
DR   INTERPRO; IPR000372; -.
DR   INTERPRO; IPR000434; -.
DR   INTERPRO; IPR000483; -.
DR   INTERPRO; IPR000601; -.
DR   INTERPRO; IPR001024; -.
DR   INTERPRO; IPR001052; -.
DR   INTERPRO; IPR001304; -.
DR   INTERPRO; IPR001611; -.
DR   INTERPRO; IPR002859; -.
DR   INTERPRO; IPR002889; -.
DR   PFAM; PF00059; lectin_c; 1.
DR   PFAM; PF00560; LRR; 2.
DR   PFAM; PF00801; PKD; 16.
DR   PFAM; PF01462; LRRNT; 1.
DR   PFAM; PF01463; LRRCT; 1.
DR   PFAM; PF01477; PLAT; 1.
DR   PFAM; PF01822; WSC; 1.
DR   PFAM; PF01825; GPS; 1.
DR   PFAM; PF02010; REJ; 1.
DR   PRINTS; PR00500; POLYCYSTIN.
DR   PROSITE; PS00202; RUBREDOXIN; UNKNOWN_1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
SQ   SEQUENCE   4292 AA;  461400 MW;  4D11600ACAA9C6AB CRC64;

  Query Match             72.9%;  Score 35;  DB 4;  Length 4292;
  Best Local Similarity   75.0%;  Pred. No. 2.7e+02;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGACLN 8
           || | |||
Db    2388 YLEGRCLN 2395


RESULT   5
Q15140
ID   Q15140      PRELIMINARY;      PRT;  4302 AA.
AC   Q15140;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   POLYCYSTIC KIDNEY DISEASE 1 PROTEIN.
GN   PKD1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94273192.
RA   Ward C.J., Peral B., Hughes J., Thomas S., Gamble V., MacCarthy A.B.,
RA   Sloane-Stanley J., Buckle V.J., Kearney L., Higgs D.R.,
RA   Ratcliffe P.J., Harris P.C., Roelfsema J.H., Spruit L.L., Saris J.J.,
RA   Dauwerse H.G., Peters D.J.M., Breuning M.H., Nellist M.,
RA   Brook-Carter P.T., Maheshwar M.M., Cordeiro I., Santos H., Cabral P.,
RA   Sampson J.R., Janssen B., Hesseling-Janssen A.L.W.,
RA   den Ouweland A.M.W., Eussen B., Verhoef S., Lindhout D.,
RA   Halley D.J.J.;
RT   "The polycystic kidney disease 1 gene encodes a 14 kb transcript and
RT   lies within a duplicated region on chromosome 16. The European
RT   Polycystic Kidney Disease Consortium [published errata appear in Cell
RT   1994 Aug 26;78(4):following 724 and 1995 Jun 30;81(7):following
RT   1170].";
RL   Cell 77:881-894(1994).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95392566.
RA   Hughes J., Ward C.J., Peral B., Aspinwall R., Clark K., Millan J.L.,
RA   Gamble V., Harris P.C.;
RT   "The polycystic kidney disease 1 (PKD1) gene encodes a novel protein
RT   with multiple cell recognition domains.";
RL   Nat. Genet. 10:151-160(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Harris P.C.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Harris P.C.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L33243; AAC37576.1; -.
DR   INTERPRO; IPR000203; -.
DR   INTERPRO; IPR000372; -.
DR   INTERPRO; IPR000434; -.
DR   INTERPRO; IPR000483; -.
DR   INTERPRO; IPR000601; -.
DR   INTERPRO; IPR001024; -.
DR   INTERPRO; IPR001052; -.
DR   INTERPRO; IPR001304; -.
DR   INTERPRO; IPR001611; -.
DR   INTERPRO; IPR002859; -.
DR   INTERPRO; IPR002889; -.
DR   PFAM; PF00059; lectin_c; 1.
DR   PFAM; PF00560; LRR; 2.
DR   PFAM; PF00801; PKD; 16.
DR   PFAM; PF01462; LRRNT; 1.
DR   PFAM; PF01463; LRRCT; 1.
DR   PFAM; PF01477; PLAT; 1.
DR   PFAM; PF01822; WSC; 1.
DR   PFAM; PF01825; GPS; 1.
DR   PFAM; PF02010; REJ; 1.
DR   PRINTS; PR00500; POLYCYSTIN.
DR   PROSITE; PS00202; RUBREDOXIN; UNKNOWN_1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
SQ   SEQUENCE   4302 AA;  462436 MW;  97472CA1981FB51A CRC64;

  Query Match            72.9%;  Score 35;  DB 4;  Length 4302;
  Best Local Similarity   75.0%;  Pred. No. 2.7e+02;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGACLN 8
           || | |||
Db    2388 YLEGRCLN 2395

RESULT   6
Q31147
ID   Q31147      PRELIMINARY;      PRT;    91 AA.
AC   Q31147;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MHC C6/G2 (QA-2) PROTEIN (FRAGMENT).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85206117.
RA   Rogers J.H.;
RT   "Family organization of mouse H-2 class I genes.";
RL   Immunogenetics 21:343-353(1985).
DR   EMBL; M14833; AAA39566.1; -.
DR   INTERPRO; IPR001039; -.
DR   PFAM; PF00129; MHC_I; 1.
KW   MHC.
```

```
FT   NON_TER       1      1
SQ   SEQUENCE   91 AA;  10747 MW;  E14021DAB93E95E8 CRC64;

  Query Match            70.8%;  Score 34;  DB 7;  Length 91;
  Best Local Similarity   85.7%;  Pred. No. 11;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGACL 7
           || ||||
Db      68 YLEGACL 74

RESULT   7
Q9SHT3
ID   Q9SHT3      PRELIMINARY;      PRT;   265 AA.
AC   Q9SHT3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   AT2G05410 PROTEIN.
GN   AT2G05410.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC   Brassicaceae; Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL   Nature 402:761-768(1999).
DR   EMBL; AC007289; AAD32896.1; -.
SQ   SEQUENCE   265 AA;  30214 MW;  C4A78A6E8CF4C0C3 CRC64;

  Query Match            70.8%;  Score 34;  DB 10;  Length 265;
  Best Local Similarity   66.7%;  Pred. No. 31;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
           ||  ||:||
Db     151 YLKTACMNL 159

RESULT   8
O75098
ID   O75098      PRELIMINARY;      PRT;   375 AA.
AC   O75098;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MEGF9 (FRAGMENT).
GN   MEGF9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 98360089.
RA   Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT   "Identification of high-molecular-weight proteins with multiple EGF-
RT   like motifs by motif-trap screening.";
RL   Genomics 51:27-34(1998).
DR   EMBL; AB011542; BAA32470.1; -.
DR   HSSP; P02468; 1TLE.
```

```
DR   INTERPRO; IPR000561; -.
DR   INTERPRO; IPR002049; -.
DR   PFAM; PF00053; laminin_EGF; 4.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; UNKNOWN_4.
FT   NON_TER       1      1
SQ   SEQUENCE   375 AA;  40803 MW;  84D1746CA8E3B172 CRC64;

  Query Match             70.8%;  Score 34;  DB 4;  Length 375;
  Best Local Similarity   85.7%;  Pred. No. 43;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGACLN 8
           |:|||||
Db      86 LTGACLN 92


RESULT   9
O09705
ID   O09705      PRELIMINARY;      PRT;  2218 AA.
AC   O09705;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   L PROTEIN.
OS   Lassa virus.
OC   Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JOSIAH;
RX   MEDLINE; 97201617.
RA   Lukashevich I.S., Djavani M., Shapiro K., Sanchez A., Ravkov E.,
RA   Nichol S.T., Salvato M.S.;
RT   "The Lassa fever virus L gene: nucleotide sequence, comparison, and
RT   precipitation of a predicted 250 kDa protein with monospecific
RT   antiserum.";
RL   J. Gen. Virol. 78:547-551(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97428590.
RA   Djavani M., Lukashevich I.S., Sanchez A., Nichol S.T., Salvato M.S.;
RT   "Completion of the Lassa fever virus sequence and identification of a
RT   RING finger open reading frame at the L RNA 5' End.";
RL   Virology 235:414-418(1997).
DR   EMBL; U63094; AAB50401.1; -.
DR   EMBL; U73034; AAC05817.1; -.
DR   INTERPRO; IPR001005; -.
DR   PROSITE; PS00334; MYB_2; UNKNOWN_1.
SQ   SEQUENCE   2218 AA;  253429 MW;  F325C51F96ED2EA6 CRC64;

  Query Match             70.8%;  Score 34;  DB 12;  Length 2218;
  Best Local Similarity   85.7%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGACLN 8
           |||:|||
Db    1158 LSGSCLN 1164


RESULT  10
Q27167
ID   Q27167      PRELIMINARY;      PRT;  2395 AA.
AC   Q27167;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   51B TYPE SURFACE PROTEIN.
OS   Paramecium tetraurelia.
OC   Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC   Paramecium.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=STOCK 51;
RX   MEDLINE; 95098630.
RA   Scott J., Leeck C., Forney J.;
RT   "Analysis of the micronuclear B type surface protein gene in
RT   Paramecium tetraurelia.";
RL   Nucleic Acids Res. 22:5079-5084(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=STOCK 51;
RA   Scott J.M., Leeck C.L., Forney J.D.;
RL   Genetics 133:189-198(1993).
DR   EMBL; U07603; AAA81947.1; -.
DR   EMBL; L04795; AAA16710.1; -.
DR   INTERPRO; IPR000950; -.
DR   INTERPRO; IPR001680; -.
DR   INTERPRO; IPR002895; -.
DR   PFAM; PF01508; Paramecium_SA; 28.
DR   PROSITE; PS00241; RECEPTOR_CYTOKINES_1; UNKNOWN_1.
DR   PROSITE; PS00678; WD_REPEATS; UNKNOWN_1.
SQ   SEQUENCE   2395 AA;  246430 MW;  8632E022649B2D61 CRC64;

  Query Match             70.8%;  Score 34;  DB 5;  Length 2395;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGAC 6
           ||||||
Db     892 YLSGAC 897


RESULT  11
O37322
ID   O37322      PRELIMINARY;      PRT;    85 AA.
AC   O37322;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JAN-1998 (TrEMBLrel. 05, Last annotation update)
DE   ORF F4L (FRAGMENT).
OS   Heliothis armigera entomopoxvirus (HaEPV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLONAL ISOLATE WT#2/011293;
RA   Sriskantha A., Osborne R.J., Dall D.J.;
RL   J. Gen. Virol. 0:0-0(1997).
DR   EMBL; AF019224; AAB96627.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   85 AA;  9920 MW;  D7A7449AD155389A CRC64;

  Query Match             68.8%;  Score 33;  DB 12;  Length 85;
  Best Local Similarity   71.4%;  Pred. No. 17;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGACLNL 9
           ||||:|:
Db      20 SGACINI 26


RESULT  12
Q22960
ID   Q22960      PRELIMINARY;      PRT;   114 AA.
AC   Q22960;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   SIMILARITY TO C4-TYPE ZINC FINGERS.
GN   F10G2.9.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Murray J., Wohldmann P.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U64836; AAB04834.1; -.
DR   INTERPRO; IPR001628; -.
DR   INTERPRO; IPR002219; -.
DR   PFAM; PF00105; zf-C4; 1.
SQ   SEQUENCE   114 AA;  13337 MW;  F9830BB98EA84E89 CRC64;


  Query Match             68.8%;  Score 33;  DB 5;  Length 114;
  Best Local Similarity   55.6%;  Pred. No. 22;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              |::| | ||
Db         47 YMAGNCFNL 55


RESULT  13
O94038
ID   O94038      PRELIMINARY;      PRT;   348 AA.
AC   O94038;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ALCOHOL DEHYDROGENASE II.
GN   CA41C10.04.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   anamorphic Saccharomycetales; Candida.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RA   Taylor K., Harris D.;
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RA   Barrell B.G., Rajandream M.A.;
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE; 97435544.
RA   Tait E., Simon M.C., King S., Brown A.J., Gow N.A.R., Shaw D.J.;
RT   "A Candida albicans genome project: cosmid contigs, physical mapping,
RT   and gene isolation.";
RL   Fungal Genet. Biol. 21:308-314(1997).
```

```
DR   EMBL; AL033501; CAA21988.1; -.
DR   INTERPRO; IPR002085; -.
DR   INTERPRO; IPR002328; -.
DR   PFAM; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc.
SQ   SEQUENCE   348 AA;  36807 MW;  3D6E67538522ADCF CRC64;

  Query Match             68.8%;  Score 33;  DB 3;  Length 348;
  Best Local Similarity   62.5%;  Pred. No. 63;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGACLN 8
              :|:|:|||
Db         93 WLNGSCLN 100


RESULT  14
Q9TP68
ID   Q9TP68      PRELIMINARY;      PRT;   362 AA.
AC   Q9TP68;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   DJ377H14.9 (MAJOR HISTOCOMPATIBILITY COMPLEX, CLASS I, F (CDA12)).
GN   HLA-F.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Williams S.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL022723; CAB46623.1; -.
DR   INTERPRO; IPR000495; -.
DR   INTERPRO; IPR001039; -.
DR   INTERPRO; IPR001064; -.
DR   INTERPRO; IPR003006; -.
DR   PFAM; PF00047; ig; 1.
DR   PFAM; PF00129; MHC_I; 1.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   MHC.
SQ   SEQUENCE   362 AA;  40578 MW;  970D072C813A07E2 CRC64;


  Query Match             68.8%;  Score 33;  DB 7;  Length 362;
  Best Local Similarity   66.7%;  Pred. No. 65;
  Matches    6;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              || | || |
Db        180 YLEGECLEL 188


RESULT  15
O74484
ID   O74484      PRELIMINARY;      PRT;   363 AA.
AC   O74484; P78779;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   PROBABLE MANNOSE-1-PHOSPHATE GUANYLTRANSFERASE (EC 2.7.7.13) (ATP-
DE   MANNOSE-1-PHOSPHATE GUANYLYLTRANSFERASE).
GN   SPCC1906.01.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC   Schizosaccharomycetaceae; Schizosaccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Lyne M., Rajandream M.A., Barrell B.G., Murphy L., Harris D.;
```

```
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PR745;
RX   MEDLINE; 98162722.
RA   Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT   "Identification of open reading frames in Schizosaccharomyces pombe
RT   cDNAs.";
RL   DNA Res. 4:363-369(1997).
CC   -!- FUNCTION: ESSENTIAL PROTEIN WHICH MAY BE INVOLVED IN THE
CC       REGULATION OF CELL CYCLE PROGRESSION (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: GTP + ALPHA-D-MANNOSE 1-PHOSPHATE =
CC       DIPHOSPHATE + GDP-MANNOSE.
CC   -!- SIMILARITY: TO S.TYPHIMURIUM CDP-GLUCOSE PYROPHOSPHORYLASE (RFBF).
DR   EMBL; AL031538; CAA20770.1; -.
DR   EMBL; D89128; BAA13790.1; -.
DR   INTERPRO; IPR001451; -.
DR   INTERPRO; IPR001825; -.
DR   PFAM; PF00132; hexapep; 1.
DR   PFAM; PF00483; NTP_transferase; 1.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
KW   Transferase; Nucleotidyltransferase; Kinase; Cell cycle.
FT   CONFLICT    175    176       GI -> VL (IN REF. 2).
SQ   SEQUENCE   363 AA;  39719 MW;  55F55FF0ED017C02 CRC64;


  Query Match             68.8%;  Score 33;  DB 3;  Length 363;
  Best Local Similarity   66.7%;  Pred. No. 65;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
             ||:| || |
Db     226 YLTGTCLYL 234


Search completed: December 16, 2000, 02:01:00
Job time: 5934 sec
```

```
                 GenCore version 4.5
          Copyright (c) 1993 - 2000  Compugen Ltd.


OM protein - protein search, using sw model

Run on:          December 16, 2000, 01:51:16 ; Search time 91.14 Seconds
                                              (without alignments)
                                              6.267 Million cell updates/sec

Title:           US-09-529-121-5
Perfect score:   48
Sequence:        1 YLSGACLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters:      182106

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_65:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                   %
Result           Query
  No.    Score   Match Length DB  ID                    Description
-----------------------------------------------------------------------
    1       36    75.0    702  2  A36319                carcinoembryonic a
    2       36    75.0   3328  2  T30835                breast cancer tumo
    3       36    75.0   3329  2  T42205                breast cancer susc
    4       36    75.0   3329  2  T30904                breast cancer tumo
    5       35    72.9     57  2  S19088                dihydrolipoamide d
    6       35    72.9    314  2  S56055                hypothetical prote
    7       35    72.9   4302  2  A38971                polycystic kidney
    8       34    70.8     91  2  I68703                MHC c6/g2 (Qa-2) p
    9       34    70.8    219  2  B49181                beta B2-crystallin
   10       34    70.8    451  2  S15236                glutathione reduct
   11       34    70.8   2395  1  S50820                surface protein ty
   12       33    68.8    114  2  T29869                hypothetical prote
   13       33    68.8    261  2  JC5806                aquaporin 8 - mous
   14       33    68.8    263  2  JC5622                aquaporin 8 - rat
   15       33    68.8    345  2  S07114                MHC class I histoc
   16       33    68.8    348  2  T18230                alcohol dehydrogen
   17       33    68.8    348  2  S29990                histocompatibility
   18       33    68.8    362  2  A60384                MHC class I histoc
   19       33    68.8    363  2  T41209                mannose-1-phosphat
   20       33    68.8    363  2  T42371                probable mannose-1
   21       33    68.8    368  2  E64142                hypothetical prote
   22       33    68.8    457  2  S73774                dihydrolipoamide d
   23       33    68.8    457  2  I64229                dihydrolipoamide d
   24       33    68.8    637  1  YCRP                  acetolactate synth
   25       33    68.8    659  2  S60058                acetolactate synth
   26       33    68.8    659  2  S60056                acetolactate synth
   27       33    68.8    670  1  YCMU                  acetolactate synth
   28       33    68.8    933  2  E64603                hypothetical prote
   29       33    68.8   1017  2  T30542                major surface glyc
   30       33    68.8   2210  1  RRXPLC                genome polyprotein
   31       32    66.7    140  2  T28576                6R protein - vario
   32       32    66.7    140  2  H36851                RNA-binding ribonu
   33       32    66.7    146  2  B81260                ribonuclease HI (E
   34       32    66.7    146  2  H72167                A35R protein - var
   35       32    66.7    249  1  CYRTB1                beta-crystallin B1
   36       32    66.7    253  2  S07264                beta-crystallin B1
   37       32    66.7    276  2  D70081                conserved hypothet
   38       32    66.7    277  2  S26008                hypothetical prote
   39       32    66.7    288  2  S44256                fructokinase (EC 2
   40       32    66.7    362  2  T38464                probable peptidyl-
   41       32    66.7    370  2  T05598                hypothetical prote
   42       32    66.7    378  2  T47407                hypothetical prote
   43       32    66.7    378  2  T33045                hypothetical prote
   44       32    66.7    420  2  T46460                hypothetical prote
   45       32    66.7    426  2  G02277                creatine transport


                                   ALIGNMENTS


RESULT   1
A36319
carcinoembryonic antigen precursor - human
N;Alternate names: CEA; meconium antigen 100
C;Species: Homo sapiens (man)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 31-Jan-2000
C;Accession: A36319; A27773; A31037; A25845; S08106; S31737; A44476; I54224; I59098;
R;Schrewe, H.; Thompson, J.; Bona, M.; Hefta, L.J.F.; Maruya, A.; Hassauer, M.; Shive
Mol. Cell. Biol. 10, 2738-2748, 1990
A;Title: Cloning of the complete gene for carcinoembryonic antigen: analysis of its p
A;Reference number: A36319; MUID:90258861
A;Accession: A36319
A;Molecule type: DNA
A;Residues: 1-702 <SCH>
A;Cross-references: GB:M17303; NID:g178676; PIDN:AAB59513.1; PID:g178677
A;Note: the authors show the codons TTA for residue 641-Phe and CAG for residue 646-T
R;Beauchemin, N.; Benchimol, S.; Cournoyer, D.; Fuks, A.; Stanners, C.P.
Mol. Cell. Biol. 7, 3221-3230, 1987
A;Title: Isolation and characterization of full-length functional cDNA clones for hum
A;Reference number: A27773; MUID:88038876
A;Accession: A27773
A;Molecule type: mRNA
A;Residues: 1-702 <BEA>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
R;Barnett, T.; Goebel, S.J.; Nothdurft, M.A.; Elting, J.J.
Genomics 3, 59-66, 1988
A;Title: Carcinoembryonic antigen family: characterization of cDNAs coding for NCA an
A;Reference number: A31037; MUID:89122014
A;Accession: A31037
A;Molecule type: mRNA
A;Residues: 1-702 <BAR>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
A;Note: the authors translated the codon GTG for residue 130 as Leu
R;Oikawa, S.; Nakazato, H.; Kosaki, G.
Biochem. Biophys. Res. Commun. 142, 511-518, 1987
A;Title: Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA
A;Reference number: A25845; MUID:87128144
A;Accession: A25845
A;Molecule type: mRNA
A;Residues: 5-702 <OIK>
A;Cross-references: GB:M15042; NID:g180198; PIDN:AAA51963.1; PID:g180199
R;Oikawa, S.
submitted to the EMBL Data Library, September 1989
A;Reference number: S08106
A;Accession: S08106
A;Molecule type: mRNA
A;Residues: 5-319,321-702 <OI2>
A;Cross-references: EMBL:X16455; NID:g29854; PIDN:CAA34474.1; PID:g825638
R;Barnett, T.
submitted to the EMBL Data Library, September 1991
A;Description: Genomic DNA sequence upstream of the translational start of the carcin
A;Reference number: S31737
```

```
A;Accession: S31737
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-141 <BA2>
A;Cross-references: EMBL:X62151
R;Khan, W.N.; Fraengsmyr, L.; Teglund, S.; Israelsson, A.; Bremer, K.; Hammarstroem, S.
Genomics 14, 384-390, 1992
A;Title: Identification of three new genes and estimation of the size of the carcinoembr
A;Reference number: A44476; MUID:93052339
A;Accession: A44476
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 35-141 <KHA>
R;Willcocks, T.C.; Craig, I.W.
Genomics 8, 492-500, 1990
A;Title: Characterization of the genomic organization of human carcinoembryonic antigen
A;Reference number: I54224; MUID:91139118
A;Accession: I54224
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-37 <RES>
A;Cross-references: GB:M60964; NID:g180215; PIDN:AAA51964.1; PID:g180217
R;Zimmermann, W.; Ortlieb, B.; Friedrich, R.; von Kleist, S.
Proc. Natl. Acad. Sci. U.S.A. 84, 2960-2964, 1987
A;Title: Isolation and characterization of cDNA clones encoding the human carcinoembryon
A;Reference number: I59098; MUID:87204247
A;Accession: I59098
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 331-702 <RE2>
A;Cross-references: GB:M16234; NID:g180240; PIDN:AAA51972.1; PID:g180241
R;Siepen, D.; Paxton, R.J.; Neumaier, M.; Shively, J.E.; Wagener, C.
Biochem. Biophys. Res. Commun. 147, 212-218, 1987
A;Title: Carcinoembryonic antigen (CEA) and two crossreacting antigens of 165 KD and 105
A;Reference number: A26831; MUID:87326349
A;Accession: A26831
A;Molecule type: protein
A;Residues: 35-64 <SIE>
R;Thomas, P.; Toth, C.A.
Biochem. Biophys. Res. Commun. 170, 391-396, 1990
A;Title: Carcinoembryonic antigen binding to Kupffer cells is via a peptide located at t
A;Reference number: A35490; MUID:90321257
A;Accession: A35490
A;Molecule type: protein
A;Residues: 'X',140-151,'X',153,'X',155-156 <THO>
A;Note: this is the amino terminal end of a fragment shown to mediate uptake by Kupffer
C;Comment: This heavily glycosylated membane protein of unknown function is a widely use
C;Comment: This protein may be processed at its C-terminus. It is anchored to the membra
C;Genetics:
A;Gene: GDB:CEA
A;Cross-references: GDB:119054; OMIM:114890
A;Map position: 19q13.2-19q13.2
A;Introns: 22/1; 142/1; 235/1; 320/1; 413/1; 498/1; 591/1; 676/1
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin
C;Keywords: blocked carboxyl end; glycoprotein; lipoprotein; membrane protein; phosphati
F;1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-678/Product: carcinoembryonic antigen #status predicted <MAT>
F;160-217/Domain: immunoglobulin homology <IMM1>
F;252-301/Domain: immunoglobulin homology <IMM2>
F;338-395/Domain: immunoglobulin homology <IMM3>
F;516-573/Domain: immunoglobulin homology <IMM4>
F;608-657/Domain: immunoglobulin homology <IMM5>
F;679-702/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F;678/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Gly) (in mature form

  Query Match             75.0%;  Score 36;  DB 2;  Length 702;
  Best Local Similarity   88.9%;  Pred. No. 27;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGACLNL 9
              ||||| |||
Db      605 YLSGANLNL 613

RESULT   2
T30835
breast cancer tumor suppressor Brca2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jan-2000
C;Accession: T30835
R;Sharan, S.K.; Bradley, A.
Genomics 40, 234-241, 1997
A;Title: Murine Brca2: sequence, map position, and expression pattern.
A;Reference number: Z20894; MUID:97237041
A;Accession: T30835
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3328 <SHA>
A;Cross-references: EMBL:U65594; NID:g1743859; PID:g1743860; PIDN:AAC23702.1
A;Experimental source: strain C57Bl/6
C;Genetics:
A;Gene: Brca2
A;Map position: 5
A;Note: expression of Brca2 detected in midgestation embryos and adult testis, thymus
C;Superfamily: breast cancer tumor suppressor BRCA2
C;Keywords: tumor suppressor

  Query Match             75.0%;  Score 36;  DB 2;  Length 3328;
  Best Local Similarity   77.8%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSGACLNL 9
              |||  ||||
Db     3011 YLSDECLNL 3019

RESULT   3
T42205
breast cancer susceptibility protein BRCA2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C;Accession: T42205
R;McAllister, K.A.; Haugen-Strano, A.; Hagevik, S.; Collins, N.K.; Brownlee, H.; Futr
submitted to the EMBL Data Library, February 1997
A;Description: Characterization of the mouse and rat homologs of the BRCA2 breast can
A;Reference number: Z22073
A;Accession: T42205
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3329 <MCA>
A;Cross-references: EMBL:U89652; NID:g2443438; PID:g2443439; PIDN:AAB71377.1
A;Experimental source: strain CD1; 129Sv; ICR Swiss
C;Genetics:
A;Gene: BRCA2
C;Superfamily: breast cancer tumor suppressor BRCA2

  Query Match             75.0%;  Score 36;  DB 2;  Length 3329;
  Best Local Similarity   77.8%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSGACLNL 9
              |||  ||||
Db     3011 YLSDECLNL 3019

RESULT   4
T30904
breast cancer tumor suppressor Brca2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jan-2000
C;Accession: T30904
R;Connor, F.; Smith, A.; Wooster, R.; Stratton, M.; Dixon, A.; Campbell, E.; Tait, T.
```

```
Hum. Mol. Genet. 6, 291-300, 1997
A;Title: Cloning, chromosomal mapping and expression pattern of the mouse Brca2 gene.
A;Reference number: Z20931; MUID:97217789
A;Accession: T30904
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3329 <CON>
A;Cross-references: EMBL:U82270; NID:g1854950; PID:g1854951; PIDN:AAB48306.1
C;Genetics:
A;Gene: Brca2
A;Map position: 5
C;Superfamily: breast cancer tumor suppressor BRCA2
C;Keywords: tumor suppressor


  Query Match             75.0%;  Score 36;  DB 2;  Length 3329;
  Best Local Similarity   77.8%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 YLSGACLNL 9
          |||  ||||
Db   3011 YLSDECLNL 3019


RESULT   5
S19088
dihydrolipoamide dehydrogenase (EC 1.8.1.4) - Enterococcus faecalis (fragment)
N;Alternate names: pyruvate dehydrogenase multienzyme complex chain E3
C;Species: Enterococcus faecalis
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 30-Jun-1997
C;Accession: S19088
R;Allen, A.G.; Perham, R.N.
FEBS Lett. 287, 206-210, 1991
A;Title: Two lipoyl domains in the dihydrolipoamide acetyltransferase chain of the pyruv
A;Reference number: S16988; MUID:91348216
A;Accession: S19088
A;Molecule type: DNA
A;Residues: 1-57 <ALL>
A;Note: the source is designated as Streptococcus faecalis
C;Genetics:
A;Gene: pdhD
C;Superfamily: dihydrolipoamide dehydrogenase; dihydrolipoamide dehydrogenase homology
C;Keywords: FAD; flavoprotein; lipoamide; NAD; oxidoreductase; redox-active disulfide; t
F;11-39/Region: beta-alpha-beta FAD nucleotide-binding fold
F;47-52/Disulfide bonds: redox-active #status predicted


  Query Match             72.9%;  Score 35;  DB 2;  Length 57;
  Best Local Similarity   55.6%;  Pred. No. 4;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 YLSGACLNL 9
          |: | ||||:
Db     42 YIGGVCLNV 50


RESULT   6
S56055
hypothetical protein YMR241w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YM9408.03
C;Species: Saccharomyces cerevisiae
C;Date: 27-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 17-Mar-2000
C;Accession: S56055
R;Gentles, S.; Bowman, S.
submitted to the EMBL Data Library, March 1995
A;Reference number: S56053
A;Accession: S56055
A;Molecule type: DNA
A;Residues: 1-314 <GEN>
A;Cross-references: EMBL:Z48756; NID:g736304; PID:g736307; GSPDB:GN00013; MIPS:YMR241w
C;Genetics:
A;Gene: MIPS:YMR241w
A;Map position: 13R
```

```
C;Superfamily: ADP,ATP carrier protein; ADP,ATP carrier protein repeat homology

  Query Match             72.9%;  Score 35;  DB 2;  Length 314;
  Best Local Similarity   87.5%;  Pred. No. 20;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 LSGACLNL 9
          | ||||||
Db     24 LLGACLNL 31


RESULT   7
A38971
polycystic kidney disease protein 1 precursor - human
C;Species: Homo sapiens (man)
C;Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 08-Oct-1999
C;Accession: A38971; A56520; A44604
R;Harris, P.C.
submitted to GenBank, May 1995
A;Reference number: A38971
A;Accession: A38971
A;Molecule type: mRNA
A;Residues: 1-4302 <HAR>
A;Cross-references: GB:L33243; NID:g904222; PIDN:AAC37576.1; PID:g904223
R;Alexandra Gluecksmann-Kuis, M.; Tayber, O.; Woolf, E.A.; Bougueleret, L.; Deng, N.;
 Z.; Torosian, S.; Zhou, J.; Reeders, S.T.; Bork, P.; Pohlschmidt, M.; Loehning, C.;
Cell 81, 289-298, 1995
A;Title: Polycystic kidney disease: the complete structure of the PKD1 gene and its p
A;Reference number: A56520; MUID:95254638
A;Accession: A56520
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-70,'E',72-137,'Q',139-252,'A',254-301,'D',303-690,'P',692-738,'R',740-7
5-1975,'V',1977-3389,'Q',3390-3980,'HV',3983-4003,'HV',4006-4302 <ALE>
A;Cross-references: GB:U24497; NID:g799334; PIDN:AAC50128.1; PID:g799335; GB:U24499
R;Ward, C.J.; Peral, B.; Hughes, J.; Thomas, S.; Gamble, V.; MacCarthy, A.B.; Sloane-
aris, J.J.; Dauwerse, H.G.; Peters, D.J.M.; Breuning, M.H.; Nellist, M.; Brook-Carter
W.; van den Ouweland, A.M.W.; Eussen, B.; Verhoef, S.; Lindhout, D.; Halley, D.J.J.
Cell 77, 881-894, 1994
A;Title: The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies wit
A;Reference number: A44604; MUID:94273192
A;Accession: A44604
A;Status: significant sequence differences
A;Molecule type: mRNA
A;Cross-references: GB:L33243
R;Ward, C.J.; Peral, B.; Hughes, J.; Thomas, S.; Gamble, V.; MacCarthy, A.B.; Sloane-
aris, J.J.; Dauwerse, H.G.; Peters, D.J.M.; Breuning, M.H.; Nellist, M.; Brook-Carter
W.; van den Ouweland, A.M.W.; Eussen, B.; Verhoef, S.; Lindhout, D.; Halley, D.J.J.
Cell 78, 724A, 1994
A;Reference number: A38972; MUID:94349375
A;Contents: annotation; erratum
A;Note: this is a revision to the sequence from reference A44604; the citation appear
R;Ward, C.J.; Peral, B.; Hughes, J.; Thomas, S.; Gamble, V.; MacCarthy, A.B.; Sloane-
aris, J.J.; Dauwerse, H.G.; Peters, D.J.M.; Breuning, M.H.; Nellist, M.; Brook-Carter
W.; van den Ouweland, A.M.W.; Eussen, B.; Verhoef, S.; Lindhout, D.; Halley, D.J.J.
Cell 81, 1171, 1995
A;Reference number: A56732
A;Contents: annotation; erratum
A;Note: this is a revision to the sequence from reference A44604
C;Genetics:
A;Gene: GDB:PKD1
A;Cross-references: GDB:120293; OMIM:173900; OMIM:601313
A;Map position: 16p13.3-16p13.3
C;Superfamily: human polycystic kidney disease protein 1; proteoglycan carboxyl-termi
C;Keywords: duplication
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-4302/Product: polycystic kidney disease protein 1 #status predicted <MAT>
F;123-170/Domain: proteoglycan carboxyl-terminal homology <PCH>


  Query Match             72.9%;  Score 35;  DB 2;  Length 4302;
  Best Local Similarity   75.0%;  Pred. No. 2.3e+02;
```

```
  Matches      6;  Conservative     0;  Mismatches     2;  Indels     0;  Gaps     0;

Qy        1 YLSGACLN 8
            || | |||
Db     2388 YLEGRCLN 2395


RESULT    8
I68703
MHC c6/g2 (Qa-2) protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 23-Jul-1999
C;Accession: I68703
R;Rogers, J.H.
Immunogenetics 21, 343-353, 1985
A;Title: Family organization of mouse H-2 class I genes.
A;Reference number: I54413; MUID:85206117
A;Accession: I68703
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-91 <RES>
A;Cross-references: GB:M14833; NID:g199298; PIDN:AAA39566.1; PID:g554214
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology


  Query Match             70.8%;  Score 34;  DB 2;  Length 91;
  Best Local Similarity   85.7%;  Pred. No. 9.8;
  Matches      6;  Conservative     0;  Mismatches     1;  Indels     0;  Gaps     0;

Qy        1 YLSGACL 7
            || ||||
Db       68 YLEGACL 74


RESULT    9
B49181
beta B2-crystallin - chicken
C;Species: Gallus gallus (chicken)
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
C;Accession: B49181
R;Sawada, K.; Agata, K.; Eguchi, G.
Exp. Eye Res. 55, 879-887, 1992
A;Title: Crystallin gene expression in the process of lentoidogenesis in cultures of chi
A;Reference number: A49181; MUID:93137981
A;Accession: B49181
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-219 <SAW>
A;Cross-references: GB:S52930; NID:g264001; PIDN:AAB25042.1; PID:g264002
A;Experimental source: lens epithelial cells
A;Note: sequence extracted from NCBI backbone (NCBIN:123033, NCBIP:123034)
C;Superfamily: beta-crystallin
C;Keywords: duplication


  Query Match             70.8%;  Score 34;  DB 2;  Length 219;
  Best Local Similarity   87.5%;  Pred. No. 22;
  Matches      7;  Conservative     0;  Mismatches     1;  Indels     0;  Gaps     0;

Qy        2 LSGACLNL 9
            ||||| ||
Db       34 LSGACPNL 41


RESULT   10
S15236
glutathione reductase (NADPH) (EC 1.6.4.2) - Pseudomonas aeruginosa
C;Species: Pseudomonas aeruginosa
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 11-Jun-1999
C;Accession: S15236
R;Perry, A.C.F.; Bhriain, N.N.; Brown, N.L.; Rouch, D.A.
Mol. Microbiol. 5, 163-171, 1991
A;Title: Molecular characterization of the gor gene encoding glutathione reductase from
```

```
ses.
A;Reference number: S15235; MUID:91194546
A;Accession: S15236
A;Molecule type: DNA
A;Residues: 1-451 <PER>
A;Cross-references: EMBL:X54201; NID:g45324; PIDN:CAA38122.1; PID:g45326
C;Superfamily: dihydrolipoamide dehydrogenase; dihydrolipoamide dehydrogenase homolog
C;Keywords: FAD; flavoprotein; NADP; oxidoreductase
F;8-443/Domain: dihydrolipoamide dehydrogenase homology <DLD>


  Query Match             70.8%;  Score 34;  DB 2;  Length 451;
  Best Local Similarity   55.6%;  Pred. No. 44;
  Matches      5;  Conservative     2;  Mismatches     2;  Indels     0;  Gaps     0;

Qy        1 YLSGACLNL 9
            || | |:|:
Db       37 YLGGTCVNV 45


RESULT   11
S50820
surface protein type 51B - Paramecium tetraurelia
C;Species: Paramecium tetraurelia
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 07-Dec-1999
C;Accession: S50820
R;Scott, J.; Leeck, C.; Forney, J.
Nucleic Acids Res. 22, 5079-5084, 1994
A;Title: Analysis of the micronuclear B type surface protein gene in Paramecium tetra
A;Reference number: S50820; MUID:95098630
A;Accession: S50820
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2395 <SCO>
A;Cross-references: EMBL:U07603; NID:g467226; PIDN:AAA81947.1; PID:g467227
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994
C;Genetics:
A;Genetic code: SGC5
A;Introns: 472/3; 1310/3; 1821/3
C;Superfamily: G surface protein


  Query Match             70.8%;  Score 34;  DB 1;  Length 2395;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches      6;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy        1 YLSGAC 6
            ||||||
Db      892 YLSGAC 897


RESULT   12
T29869
hypothetical protein F10G2.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T29869
R;Murray, J.; Wohldmann, P.
submitted to the EMBL Data Library, July 1996
A;Description: The sequence of C. elegans cosmid F10G2.
A;Reference number: Z20701
A;Accession: T29869
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-114 <MUR>
A;Cross-references: EMBL:U64836; PIDN:AAB04834.1; GSPDB:GN00023; CESP:F10G2.9
A;Experimental source: strain Bristol N2; clone F10G2
C;Genetics:
A;Gene: CESP:F10G2.9
A;Map position: 5
A;Introns: 61/3; 94/3
```

```
  Query Match             68.8%;  Score 33;  DB 2;  Length 114;
  Best Local Similarity   55.6%;  Pred. No. 19;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
           |::| | ||
Db      47 YMAGNCFNL 55


RESULT  13
JC5806
aquaporin 8 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 27-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 13-Aug-1999
C;Accession: JC5806
R;Ma, T.; Yang, B.; Verkman, A.S.
Biochem. Biophys. Res. Commun. 240, 324-328, 1997
A;Title: Cloning of a novel water and urea-permeable aquaporin from mouse expressed strc
A;Reference number: JC5806; MUID:98049830
A;Accession: JC5806
A;Molecule type: mRNA
A;Residues: 1-261 <MAA>
A;Cross-references: DDBJ:AF018952; NID:g2353796; PIDN:AAB68847.1; PID:g2353797
C;Comment: This protein functions as a mercurial-sensitive water channel.
C;Superfamily: lens fiber membrane major intrinsic protein
C;Keywords: glycoprotein
F;92-94/Region: NPA motif
F;210-212/Region: NPA motif
F;85,139/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             68.8%;  Score 33;  DB 2;  Length 261;
  Best Local Similarity   71.4%;  Pred. No. 41;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGACLN 8
           :||||:|
Db     204 ISGACMN 210


RESULT  14
JC5622
aquaporin 8 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 09-Oct-1997 #sequence_revision 07-Nov-1997 #text_change 20-Jun-2000
C;Accession: JC5622
R;Ishibashi, K.; Kuwahara, M.; Kageyama, Y.; Tohsaka, A.; Marumo, F.; Sasaki, S.
Biochem. Biophys. Res. Commun. 237, 714-718, 1997
A;Title: Cloning and functional expression of a second new aquaporin abundantly expresse
A;Reference number: JC5622; MUID:97445104
A;Contents: Testis
A;Accession: JC5622
A;Molecule type: mRNA
A;Residues: 1-263 <ISH>
A;Cross-references: DDBJ:AB005547; NID:g2346967; PIDN:BAA21918.1; PID:g2346968
C;Comment: This protein is a water channel protein which plays a role in the regulation
C;Superfamily: lens fiber membrane major intrinsic protein
C;Keywords: glycoprotein
F;39-59/Domain: transmembrane #status predicted <TM1>
F;65-84/Domain: transmembrane #status predicted <TM2>
F;94-96/Region: NPA motif
F;109-130/Domain: transmembrane #status predicted <TM3>
F;159-179/Domain: transmembrane #status predicted <TM4>
F;182-204/Domain: transmembrane #status predicted <TM5>
F;212-214/Region: NPA motif
F;231-250/Domain: transmembrane #status predicted <TM6>
F;141/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             68.8%;  Score 33;  DB 2;  Length 263;
  Best Local Similarity   71.4%;  Pred. No. 42;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGACLN 8
           :||||:|
Db     206 ISGACMN 212


RESULT  15
S07114
MHC class I histocompatibility antigen Ch28 alpha chain precursor - chimpanzee
C;Species: Pan troglodytes (chimpanzee)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jan-2000
C;Accession: S07114
R;Lawlor, D.A.; Ward, F.E.; Ennis, P.D.; Jackson, A.P.; Parham, P.
Nature 335, 268-271, 1988
A;Title: HLA-A and B polymorphisms predate the divergence of humans and chimpanzees.
A;Reference number: S06424; MUID:88319000
A;Accession: S07114
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-345 <LAW>
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology
C;Keywords: glycoprotein; membrane protein
F;217-282/Domain: immunoglobulin homology <IMM>


  Query Match             68.8%;  Score 33;  DB 2;  Length 345;
  Best Local Similarity   66.7%;  Pred. No. 54;
  Matches    6;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
           || | || |
Db     180 YLEGECLEL 188


Search completed: December 16, 2000, 01:51:17
Job time: 7638 sec
```

```
                         GenCore version 4.5
                Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 03:05:20 ; Search time 69.45 Seconds
                                               (without alignments)
                                               4.140 Million cell updates/sec

Title:           US-09-529-121-5
Perfect score:   48
Sequence:        1 YLSGACLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters:      87993

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_39:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                                   SUMMARIES
                     %
Result              Query
  No.     Score     Match Length DB  ID                         Description
-----------------------------------------------------------------------------
    1        36     75.0    702  1  CCEM_HUMAN                 P06731 homo sapien
    2        36     75.0   3329  1  BRC2_MOUSE                 P97929 mus musculu
    3        35     72.9    314  1  YM78_YEAST                 Q04013 saccharomyc
    4        35     72.9   4303  1  PKD1_HUMAN                 P98161 homo sapien
    5        34     70.8    218  1  CRB2_CHICK                 Q05714 gallus gall
    6        34     70.8    451  1  GSHR_PSEAE                 P23189 pseudomonas
    7        33     68.8    261  1  AQP8_MOUSE                 P56404 mus musculu
    8        33     68.8    263  1  AQP8_RAT                   P56405 rattus norv
    9        33     68.8    346  1  1C28_PANTR                 P16215 pan troglod
   10        33     68.8    348  1  ADH1_PICST                 O00097 pichia stip
   11        33     68.8    348  1  HLAF_MACMU                 P33617 macaca mula
   12        33     68.8    362  1  HLAF_HUMAN                 P30511 homo sapien
   13        33     68.8    368  1  Y093_HAEIN                 P44509 haemophilus
   14        33     68.8    457  1  DLDH_MYCGE                 P47513 mycoplasma
   15        33     68.8    457  1  DLDH_MYCPN                 P75393 mycoplasma
   16        33     68.8    637  1  ILV2_BRANA                 P14874 brassica na
   17        33     68.8    670  1  ILVB_ARATH                 P17597 arabidopsis
   18        33     68.8   2210  1  RRPO_LYCVA                 P14240 lymphocytic
   19        32     66.7    140  1  VA31_VARV                  P33848 variola vir
   20        32     66.7    248  1  CRB1_RAT                   P02523 rattus norv
   21        32     66.7    252  1  CRB1_BOVIN                 P07318 bos taurus
   22        32     66.7    276  1  YXKO_BACSU                 P94368 bacillus su
   23        32     66.7    277  1  YM06_MARPO                 P38454 marchantia
   24        32     66.7    288  1  SCRK_PEDPE                 P43468 pediococcus
   25        32     66.7    362  1  YAV6_SCHPO                 Q10175 schizosacch
   26        32     66.7    426  1  NTCS_HUMAN                 P53796 homo sapien
   27        32     66.7    440  1  GAAP_RAT                   O09028 rattus norv
   28        32     66.7    510  1  NOA1_HUMAN                 P51513 homo sapien
   29        32     66.7    607  1  ETFD_PIG                   P55931 sus scrofa
   30        32     66.7    617  1  ETFD_HUMAN                 Q16134 homo sapien
   31        32     66.7    667  1  TSPE_BPP22                 P12528 bacteriopha
   32        32     66.7    898  1  NAH5_RAT                   Q9z0x2 rattus norv
   33        32     66.7    954  1  M3KA_HUMAN                 Q02779 homo sapien
   34        32     66.7   4540  1  DYHC_PARTE                 Q27171 paramecium
   35      31.5     65.6    536  1  VNCA_AAV2                  P03132 adeno-assoc
   36        31     64.6     46  1  AX2_BETVU                  P82010 beta vulgar
   37        31     64.6    326  1  HA18_MOUSE                 P14430 mus musculu
   38        31     64.6    328  1  HA1Q_MOUSE                 P14428 mus musculu
   39        31     64.6    368  1  HA14_MOUSE                 P14427 mus musculu
   40        31     64.6    368  1  HA1W_MOUSE                 P03991 mus musculu
   41        31     64.6    459  1  DLD1_PSEPU                 P09063 pseudomonas
   42        31     64.6    511  1  IE63_HSV1E                 P36295 herpes simp
   43        31     64.6    512  1  IE63_HSV11                 P10238 herpes simp
   44        31     64.6    599  1  VG58_HSVI1                 Q00157 ictalurid h
   45        31     64.6    662  1  I12R_HUMAN                 P42701 homo sapien

                                   ALIGNMENTS

RESULT   1
CCEM_HUMAN
ID   CCEM_HUMAN     STANDARD;      PRT;   702 AA.
AC   P06731;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINOEMBRYONIC ANTIGEN PRECURSOR (CEA) (MECONIUM ANTIGEN 100)
DE   (CD66E ANTIGEN).
GN   CEA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90258861.
RA   Schrewe H., Thompson J., Bona M., Hefta L.J.F., Maruya A.,
RA   Hassauer M., Shively J.E., von Kleist S., Zimmermann W.;
RT   "Cloning of the complete gene for carcinoembryonic antigen: analysis
RT   of its promoter indicates a region conveying cell type-specific
RT   expression.";
RL   Mol. Cell. Biol. 10:2738-2748(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88038876.
RA   Beauchemin N., Benchimol S., Cournoyer D., Fuks A., Stanners C.P.;
RT   "Isolation and characterization of full-length functional cDNA clones
RT   for human carcinoembryonic antigen.";
RL   Mol. Cell. Biol. 7:3221-3230(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89122014.
RA   Barnett T., Goebel S.J., Nothdurft M.A., Elting J.J.;
RT   "Carcinoembryonic antigen family: characterization of cDNAs coding
RT   for NCA and CEA and suggestion of nonrandom sequence variation in
RT   their conserved loop-domains.";
RL   Genomics 3:59-66(1988).
RN   [4]
RP   SEQUENCE OF 5-702 FROM N.A.
RX   MEDLINE; 87128144.
RA   Oikawa S., Nakazato H., Kosaki G.;
RT   "Primary structure of human carcinoembryonic antigen (CEA) deduced
RT   from cDNA sequence.";
RL   Biochem. Biophys. Res. Commun. 142:511-518(1987).
RN   [5]
RP   SEQUENCE OF 331-702 FROM N.A.
RX   MEDLINE; 87204247.
RA   Zimmermann W., Ortlieb B., Friedrich R., von Kleist S.;
RT   "Isolation and characterization of cDNA clones encoding the human
RT   carcinoembryonic antigen reveal a highly conserved repeating
RT   structure.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:2960-2964(1987).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC   -!- TISSUE SPECIFICITY: FOUND IN ADENOCARCINOMAS OF ENDODERMALLY
CC       DERIVED DIGESTIVE SYSTEM EPITHELIUM AND FETAL COLON.
CC   -!- PTM: COMPLEX IMMUNOREACTIVE GLYCOPROTEIN WITH A MW OF 180 KDA
```

```
CC       COMPRISING 60% CARBOHYDRATE.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       7 IG-LIKE DOMAINS. BELONGS TO THE CARCINOEMBRYONIC ANTIGEN
CC       SUBFAMILY.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD66e entry;
CC       WWW="HTTP://WWW.NCBI.NLM.NIH.GOV/PROW/CD/CD66E.HTM".
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M17303; AAB59513.1; -.
DR   EMBL; M59262; AAA62835.1; ALT_SEQ.
DR   EMBL; M59255; AAA62835.1; JOINED.
DR   EMBL; M59256; AAA62835.1; JOINED.
DR   EMBL; M59257; AAA62835.1; JOINED.
DR   EMBL; M59258; AAA62835.1; JOINED.
DR   EMBL; M59259; AAA62835.1; JOINED.
DR   EMBL; M59260; AAA62835.1; JOINED.
DR   EMBL; M59261; AAA62835.1; JOINED.
DR   EMBL; M59709; -; NOT_ANNOTATED_CDS.
DR   EMBL; M59710; -; NOT_ANNOTATED_CDS.
DR   EMBL; M29540; AAA51967.1; -.
DR   EMBL; X16455; CAA34474.1; -.
DR   EMBL; M15042; AAA51963.1; -.
DR   EMBL; M16234; AAA51972.1; -.
DR   PIR; A36319; A36319.
DR   MIM; 114890; -.
DR   INTERPRO; IPR003006; -.
DR   PFAM; PF00047; ig; 7.
KW   Immunoglobulin domain; Glycoprotein; GPI-anchor; Membrane; Signal.
FT   SIGNAL        1     34
FT   CHAIN        35      ?       CARCINOEMBRYONIC ANTIGEN.
FT   PROPEP        ?    702       REMOVED IN MATURE FORM.
FT   DOMAIN       35    144       IG-LIKE DOMAIN 1.
FT   DOMAIN      146    237       IG-LIKE DOMAIN 2.
FT   DOMAIN      238    322       IG-LIKE DOMAIN 3.
FT   DOMAIN      324    415       IG-LIKE DOMAIN 4.
FT   DOMAIN      416    498       IG-LIKE DOMAIN 5.
FT   DOMAIN      502    593       IG-LIKE DOMAIN 6.
FT   DOMAIN      594    677       IG-LIKE DOMAIN 7.
FT   CARBOHYD    104    104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    115    115       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    152    152       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    182    182       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    197    197       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    204    204       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    208    208       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    246    246       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    256    256       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    274    274       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    288    288       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    292    292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    309    309       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    330    330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    351    351       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    360    360       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    375    375       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    432    432       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    466    466       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    480    480       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    508    508       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    529    529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    553    553       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    560    560       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    580    580       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    612    612       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    650    650       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    665    665       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CONFLICT    320    320       MISSING (IN REF. 4).
SQ   SEQUENCE   702 AA;  76795 MW;  6299AE26CDDBDB5C CRC64;

  Query Match             75.0%;  Score 36;  DB 1;  Length 702;
  Best Local Similarity   88.9%;  Pred. No. 17;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGACLNL 9
            ||||| |||
Db      605 YLSGANLNL 613

RESULT   2
BRC2_MOUSE
ID   BRC2_MOUSE     STANDARD;      PRT;  3329 AA.
AC   P97929; P97383; O35922;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   BREAST CANCER TYPE 2 SUSCEPTIBILITY PROTEIN.
GN   BRCA2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129;
RX   MEDLINE; 97217789.
RA   Connor F., Smith A., Wooster R., Stratton M., Dixon A., Campbell E.,
RA   Tait T.M., Freeman T., Ashworth A.;
RT   "Cloning, chromosomal mapping and expression pattern of the mouse
RT   Brca2 gene.";
RL   Hum. Mol. Genet. 6:291-300(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE; 97237041.
RA   Sharan S.K., Bradley A.;
RT   "Murine Brca2: sequence, map position, and expression pattern.";
RL   Genomics 40:234-241(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SV;
RX   MEDLINE; 97384941.
RA   McAllister K.A., Haugen-Strano A., Hagevik S., Brownlee H.A.,
RA   Collins N.K., Futreal P.A., Bennett L.M., Wiseman R.W.;
RT   "Characterization of the rat and mouse homologues of the BRCA2 breast
RT   cancer susceptibility gene.";
RL   Cancer Res. 57:3121-3125(1997).
RN   [4]
RP   SEQUENCE OF 18-200 FROM N.A.
RX   MEDLINE; 97075121.
RA   Rajan J.V., Wang M., Marquis S.T., Chodosh L.A.;
RT   "Brca2 is coordinately regulated with Brca1 during proliferation and
RT   differentiation in mammary epithelial cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:13078-13083(1996).
RN   [5]
RP   SEQUENCE OF 569-625 FROM N.A.
RX   MEDLINE; 97341126.
RA   McAllister K.A., Ramachandran S., Haugen-Strano A., Fiedorek F.T. Jr.,
RA   Wiseman R.W.;
RT   "Genetic mapping of the Brca2 breast cancer susceptibility gene on
RT   mouse chromosome 5.";
RL   Mamm. Genome 8:540-541(1997).
CC   -!- FUNCTION: MAY PARTICIPATE IN A PATHWAY ASSOCIATED WITH THE
CC       ACTIVATION OF DOUBLE-STRAND BREAK REPAIR AND/OR HOMOLOGOUS
CC       RECOMBINATION (BY SIMILARITY).
CC   -!- SUBUNIT: INTERACTS WITH RAD51 (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST EXPRESSION IN
CC       CEREBELLUM, TESTIS, ILEUM, APPENDIX, EPIDYMIS, OVARY AND MAMMARY
CC       GLAND. NO EXPRESSION IN LUNG.
CC   -!- DEVELOPMENTAL STAGE: IN THE MAMMARY GLAND, EXPRESSION INCREASES
```

```
CC      DRAMATICALLY DURING PREGNANCY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U82270; AAB48306.1; -.
DR   EMBL; U72947; AAB40720.1; -.
DR   EMBL; U65594; AAC23702.1; -.
DR   EMBL; U89652; AAB71377.1; -.
DR   EMBL; U89503; AAC53276.1; -.
DR   MGD; MGI:109337; BRCA2.
DR   INTERPRO; IPR002093; -.
DR   PFAM; PF00634; BRCA2_repeat; 7.
DR   PROSITE; PS50138; BRCA2_REPEAT; 6.
KW   Polymorphism; Repeat.
FT   DOMAIN      981   2038       7 X 34 AA APPROXIMATE BRCA2 REPEATS.
FT   REPEAT      981   1015       BRCA2 1.
FT   REPEAT     1192   1226       BRCA2 2.
FT   REPEAT     1394   1428       BRCA2 3.
FT   REPEAT     1491   1525       BRCA2 4.
FT   REPEAT     1623   1657       BRCA2 5.
FT   REPEAT     1924   1958       BRCA2 6.
FT   REPEAT     2004   2038       BRCA2 7.
FT   VARIANT      44     44       S -> F (IN STRAINS C57BL/6 AND 129/SV).
FT   VARIANT     340    340       T -> P (IN STRAIN 129/SV).
FT   VARIANT     377    377       N -> H (IN STRAIN C57BL/6).
FT   VARIANT     407    407       H -> P (IN STRAIN C57BL/6).
FT   VARIANT     661    661       I -> V (IN STRAIN C57BL/6).
FT   VARIANT     739    739       P -> H (IN STRAIN C57BL/6).
FT   VARIANT    1038   1038       I -> L (IN STRAINS C57BL/6 AND 129/SV).
FT   VARIANT    1198   1199       GF -> RI (IN STRAIN C57BL/6).
FT   VARIANT    1257   1257       Q -> P (IN STRAIN C57BL/6).
FT   VARIANT    1392   1392       Q -> R (IN STRAIN C57BL/6).
FT   VARIANT    1520   1521       FD -> CG (IN STRAIN C57BL/6).
FT   VARIANT    1583   1583       R -> W (IN STRAIN C57BL/6).
FT   VARIANT    1613   1613       C -> W (IN STRAIN C57BL/6).
FT   VARIANT    1686   1686       S -> R (IN STRAIN C57BL/6).
FT   VARIANT    1799   1799       S -> F (IN STRAIN 129/SV).
FT   VARIANT    1881   1881       P -> L (IN STRAIN C57BL/6).
FT   VARIANT    1894   1894       S -> F (IN STRAIN 129/SV).
FT   VARIANT    2141   2141       Q -> K (IN STRAIN C57BL/6).
FT   VARIANT    2392   2392       S -> R (IN STRAIN C57BL/6).
FT   VARIANT    2605   2605       K -> Q (IN STRAIN C57BL/6).
FT   VARIANT    2648   2648       A -> P (IN STRAIN C57BL/6).
FT   VARIANT    2717   2717       R -> C (IN STRAIN 129/SV).
FT   VARIANT    2729   2729       L -> M (IN STRAIN 129/SV).
FT   VARIANT    2814   2814       Q -> H (IN STRAIN C57BL/6).
FT   VARIANT    2827   2827       A -> P (IN STRAIN C57BL/6).
FT   VARIANT    2907   2907       S -> I (IN STRAIN 129/SV).
FT   VARIANT    2929   2929       H -> L (IN STRAIN 129/SV).
FT   VARIANT    3058   3058       A -> G (IN STRAIN C57BL/6).
FT   VARIANT    3071   3071       A -> G (IN STRAIN C57BL/6).
FT   VARIANT    3081   3081       K -> E (IN STRAIN C57BL/6).
FT   VARIANT    3089   3089       T -> S (IN STRAIN C57BL/6).
FT   VARIANT    3105   3109       DSPKW -> SQSQV (IN STRAIN C57BL/6).
FT   VARIANT    3220   3220       A -> G (IN STRAIN 129/SV).
FT   VARIANT    3238   3238       E -> K (IN STRAIN 129/SV).
FT   VARIANT    3242   3243       PR -> P (IN STRAIN C57BL/6).
FT   VARIANT    3245   3245       R -> K (IN STRAIN 129/SV).
SQ   SEQUENCE   3329 AA;  370658 MW;  A3B269AF8F05A5D1 CRC64;

  Query Match             75.0%;  Score 36;  DB 1;  Length 3329;
  Best Local Similarity   77.8%;  Pred. No. 69;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSGACLNL 9
            |||  ||||
Db     3011 YLSDECLNL 3019


RESULT   3
YM78_YEAST
ID   YM78_YEAST     STANDARD;      PRT;   314 AA.
AC   Q04013;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL 34.2 KDA PROTEIN IN CUS1-RPL20A INTERGENIC REGION.
GN   YMR241W OR YM9408.03.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Gentles S., Bowman S., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO S.POMBE SPBC83.13.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z48756; CAA88651.1; -.
DR   SGD; S0004854; YMR241W.
DR   INTERPRO; IPR001993; -.
DR   PFAM; PF00153; mito_carr; 2.
KW   Hypothetical protein.
SQ   SEQUENCE   314 AA;  34184 MW;  EB483C715D195BED CRC64;

  Query Match             72.9%;  Score 35;  DB 1;  Length 314;
  Best Local Similarity   87.5%;  Pred. No. 13;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        2 LSGACLNL 9
              | ||||||
Db       24 LLGACLNL 31


RESULT   4
PKD1_HUMAN
ID   PKD1_HUMAN     STANDARD;      PRT;  4303 AA.
AC   P98161;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   POLYCYSTIN PRECURSOR (AUTOSOMAL DOMINANT POLYCYSTIC KIDNEY DISEASE
DE   PROTEIN 1).
GN   PKD1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95254638.
RA   Gluecksmann-Kuis M.A., Tayber O., Woolf E.A., Bougueleret L.,
RA   Deng N., Alperin G.D., Iris F., Hawkins F., Munro C., Lakey N.,
RA   Duyk G., Schneider M.C., Geng L., Zhang F., Zhao Z., Torosian S.,
RA   Reeders S.T., Bork P., Pohlschmidt M., Loehning C., Kraus B.,
RA   Nowicka U., Leung A.L.S., Frischauf A.-M.;
RT   "Polycystic kidney disease: the complete structure of the PKD1 gene
RT   and its protein. The International Polycystic Kidney Disease
RT   Consortium.";
RL   Cell 81:289-298(1995).
RN   [2]
```

```
RP   SEQUENCE OF 2769-4303 FROM N.A.
RX   MEDLINE; 94273192.
RA   Ward C.J., Peral B., Hughes J., Thomas S., Gamble V.,
RA   Maccarthy A.B., Sloane-Stanley J., Buckle V.J., Kearney L.,
RA   Higgs D.R., Ratcliffe P.J., Harris P.C., Roelfsema J.H.,
RA   Spruit L.L., Saris J.J., Dauwerse H.G., Peters D.J.M.,
RA   Breuning M.H., Nellist M., Brook-Carter P.T., Maheshwar M.M.,
RA   Cordeiro I., Santos H., Cabral P., Sampson J.R., Janssen B.,
RA   Hesseling-Janssen A.L.W., van den Ouweland A.M.W., Eussen B.,
RA   Verhoef S., Lindhout D., Halley D.J.J.;
RT   "The polycystic kidney disease 1 gene encodes a 14 kb transcript and
RT   lies within a duplicated region on chromosome 16. The European
RT   Polycystic Kidney Disease Consortium.";
RL   Cell 77:881-894(1994).
RN   [3]
RP   STRUCTURE BY NMR OF 275-354.
RX   MEDLINE; 99107746.
RA   Bycroft M., Bateman A., Clarke J., Hamill S.J., Sandford R.,
RA   Thomas R.L., Chothia C.;
RT   "The structure of a PKD domain from polycystin-1: implications for
RT   polycystic kidney disease.";
RL   EMBO J. 18:297-305(1999).
RN   [4]
RP   VARIANT ADPKD 3748-ARG-VAL-3752 DEL, AND VARIANT ASP-3632.
RX   MEDLINE; 96108969.
RA   Peral B., San Millan J.L., Ong A.C.M., Gamble V., Ward C.J.,
RA   Strong C., Harris P.C.;
RT   "Screening the 3' region of the polycystic kidney disease 1 (PKD1)
RT   gene reveals six novel mutations.";
RL   Am. J. Hum. Genet. 58:86-96(1996).
RN   [5]
RP   VARIANT ALA-4058.
RX   MEDLINE; 97295081.
RA   Constantinides R., Xenophontos S., Neophytou P., Nomura S.,
RA   Pierides A., Constantinou-Deltas C.D.;
RT   "New amino acid polymorphism, Ala/Val4058, in exon 45 of the
RT   polycystic kidney disease 1 gene: evolution of alleles.";
RL   Hum. Genet. 99:644-647(1997).
RN   [6]
RP   VARIANTS T-2760;P-2761;V-2763;T-2764;Q-2791;T-2826;L-3008 & L-3064.
RX   MEDLINE; 97449169.
RA   Watnick T.J., Piontek K.B., Cordal T.M., Weber H., Gandolph M.A.,
RA   Qian F., Lens X.M., Neumann H.P.H., Germino G.G.;
RT   "An unusual pattern of mutation in the duplicated portion of PKD1 is
RT   revealed by use of a novel strategy for mutation detection.";
RL   Hum. Mol. Genet. 6:1473-1481(1997).
RN   [7]
RP   VARIANT ADPKD THR-3678.
RX   MEDLINE; 97403939.
RA   Turco A.E., Rossetti S., Bresin E., Englisch S., Corra S.,
RA   Pignatti P.F.;
RT   "Three novel mutations of the PKD1 gene in Italian families with
RT   autosomal dominant polycystic kidney disease.";
RL   Hum. Mutat. 10:164-167(1997).
RN   [8]
RP   VARIANT ADPKD ASP-4032, AND VARIANT VAL-4045.
RX   MEDLINE; 98180892.
RA   Daniells C., Maheshwar M., Lazarou L., Davies F., Coles G., Ravine D.;
RT   "Novel and recurrent mutations in the PKD1 (polycystic kidney
RT   disease) gene.";
RL   Hum. Genet. 102:216-220(1998).
RN   [9]
RP   VARIANT ADPKD MET-3375.
RX   MEDLINE; 99118881.
RA   Koptides M., Constantinides R., Kyriakides G., Hadjigavriel M.,
RA   Patsalis P.C., Pierides A., Deltas C.C.;
RT   "Loss of heterozygosity in polycystic kidney disease with a missense
RT   mutation in the repeated region of PKD1.";
RL   Hum. Genet. 103:709-717(1998).
CC   -!- FUNCTION: PKD1 AND PKD2 MAY FUNCTION THROUGH A COMMON SIGNALING
CC       PATHWAY THAT IS NECESSARY FOR NORMAL TUBULOGENESIS. INVOLVED IN
CC       ADHESIVE PROTEIN-PROTEIN AND PROTEIN-CARBOHYDRATE INTERACTIONS.
CC   -!- SUBUNIT: INTERACTS WITH PKD2.
CC   -!- DISEASE: DEFECTS IN PKD1 ARE THE CAUSE OF AUTOSOMAL DOMINANT
CC       POLYCYSTIC KIDNEY DISEASE (ADPKD), A COMMON AUTOSOMAL DOMINANT
CC       GENETIC DISEASE AFFECTING ABOUT 1 OUT 1000 INDIVIDUALS. IT IS
CC       CHARACTERIZED BY PROGRESSIVE FORMATION AND ENLARGEMENT OF CYSTS IN
CC       BOTH KIDNEYS, TYPICALLY LEADING TO END-STAGE RENAL DISEASE IN
CC       ADULT LIFE. CYSTS ALSO OCCURS IN THE LIVER AND OTHER ORGANS.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 2, FRAMED BY A
CC       LRR N-FLANK AND LRR C-FLANK.
CC   -!- SIMILARITY: CONTAINS 16 POLYCYSTIC KIDNEY DISEASE DOMAINS (PKD).
CC   -!- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN (ATYPICAL,
CC       THE POTENTIAL CALCIUM-BINDING SITE IS MISSING).
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U24497; AAC50128.1; -.
DR   PDB; 1B4R; 06-JAN-99.
DR   MIM; 601313; -.
DR   MIM; 173900; -.
DR   INTERPRO; IPR000203; -.
DR   INTERPRO; IPR000372; -.
DR   INTERPRO; IPR000434; -.
DR   INTERPRO; IPR000483; -.
DR   INTERPRO; IPR000601; -.
DR   INTERPRO; IPR001024; -.
DR   INTERPRO; IPR001304; -.
DR   INTERPRO; IPR001611; -.
DR   INTERPRO; IPR002859; -.
DR   INTERPRO; IPR002889; -.
DR   PFAM; PF01825; GPS; 1.
DR   PFAM; PF00560; LRR; 2.
DR   PFAM; PF01463; LRRCT; 1.
DR   PFAM; PF01462; LRRNT; 1.
DR   PFAM; PF00801; PKD; 17.
DR   PFAM; PF01477; PLAT; 1.
DR   PFAM; PF02010; REJ; 1.
DR   PFAM; PF01822; WSC; 1.
DR   PFAM; PF00059; lectin_c; 1.
DR   PRINTS; PR00500; POLYCYSTIN.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Signal; Leucine-repeat; Lectin; Repeat; Glycoprotein; Transmembrane;
KW   Disease mutation; Polymorphism; 3D-structure.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24   4303       POLYCYSTIN.
FT   DOMAIN      638    671       LDL-RECEPTOR CLASS A.
FT   DOMAIN       32     65       LRR N-FLANK.
FT   DOMAIN       72    119       LEUCINE-RICH REPEATS.
FT   REPEAT       72     94       LRR 1.
FT   REPEAT       97    119       LRR 2.
FT   DOMAIN      123    177       LRR C-FLANK.
FT   REPEAT      282    353       PKD 1.
FT   DOMAIN      405    534       C-TYPE LECTIN.
FT   REPEAT      844    929       PKD 2.
FT   REPEAT      930   1014       PKD 3.
FT   DOMAIN     1032   2142       13 X 80 AA REPEATS.
FT   REPEAT     1032   1124       PKD 4.
FT   REPEAT     1138   1209       PKD 5.
FT   REPEAT     1221   1292       PKD 6.
FT   REPEAT     1305   1377       PKD 7.
FT   REPEAT     1390   1463       PKD 8.
FT   REPEAT     1477   1545       PKD 9.
FT   REPEAT     1559   1629       PKD 10.
FT   REPEAT     1643   1715       PKD 11.
FT   REPEAT     1729   1799       PKD 12.
FT   REPEAT     1815   1883       PKD 13.
FT   REPEAT     1898   1968       PKD 14.
```

```
FT   REPEAT     1983   2058       PKD 15.
FT   REPEAT     2071   2142       PKD 16.
FT   TRANSMEM   2216   2236       POTENTIAL.
FT   TRANSMEM   2546   2566       POTENTIAL.
FT   TRANSMEM   2580   2600       POTENTIAL.
FT   TRANSMEM   2693   2713       POTENTIAL.
FT   TRANSMEM   2808   2828       POTENTIAL.
FT   TRANSMEM   3075   3095       POTENTIAL.
FT   TRANSMEM   3283   3303       POTENTIAL.
FT   TRANSMEM   3324   3344       POTENTIAL.
FT   TRANSMEM   3560   3580       POTENTIAL.
FT   TRANSMEM   3583   3603       POTENTIAL.
FT   TRANSMEM   3674   3694       POTENTIAL.
FT   TRANSMEM   3897   3917       POTENTIAL.
FT   TRANSMEM   3939   3959       POTENTIAL.
FT   TRANSMEM   3980   4000       POTENTIAL.
FT   TRANSMEM   4028   4048       POTENTIAL.
FT   TRANSMEM   4055   4075       POTENTIAL.
FT   TRANSMEM   4086   4106       POTENTIAL.
FT   DISULFID    640    653       BY SIMILARITY.
FT   DISULFID    647    665       BY SIMILARITY.
FT   DISULFID    660    669       BY SIMILARITY.
FT   CARBOHYD     50     50       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     89     89       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    116    116       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    121    121       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    621    621       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    632    632       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    746    746       N-LINKED (GLCNAC. . .) (POTENTIAL).

  Query Match             72.9%;  Score 35;  DB 1;  Length 4303;
  Best Local Similarity   75.0%;  Pred. No. 1.3e+02;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 YLSGACLN 8
          || | |||
Db   2388 YLEGRCLN 2395


RESULT   5
CRB2_CHICK
ID   CRB2_CHICK     STANDARD;      PRT;   218 AA.
AC   Q05714;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   BETA CRYSTALLIN B2 (BP).
GN   CRYBB2.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LENS;
RX   MEDLINE; 93137981.
RA   Sawada K., Agata K., Eguchi G.;
RT   "Crystallin gene expression in the process of lentoidogenesis in
RT   cultures of chicken lens epithelial cells.";
RL   Exp. Eye Res. 55:879-887(1992).
CC   -!- FUNCTION: CRYSTALLINS ARE THE DOMINANT STRUCTURAL COMPONENTS
CC       OF THE VERTEBRATE EYE LENS.
CC   -!- SUBUNIT: HOMO/HETERODIMER, OR COMPLEXES OF HIGHER ORDER. THE
CC       STRUCTURE OF BETA-CRYSTALLIN OLIGOMERS SEEMS TO BE STABILIZED
CC       THROUGH INTERACTIONS BETWEEN THE N-TERMINAL ARMS (BY SIMILARITY).
CC   -!- DOMAIN: HAS A TWO-DOMAIN BETA-STRUCTURE, FOLDED INTO FOUR
CC       VERY SIMILAR GREEK KEY MOTIFS.
CC   -!- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
```

```
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; S52930; AAB25042.1; -.
DR   PIR; B49181; B49181.
DR   HSSP; P02522; 1BLB.
DR   INTERPRO; IPR001064; -.
DR   PFAM; PF00030; crystall; 2.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 3.
KW   Eye lens protein; Duplication; Acetylation.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
FT   DOMAIN        1     15       N-TERMINAL ARM.
FT   DOMAIN       16     55       MOTIF 1.
FT   DOMAIN       56    101       MOTIF 2.
FT   DOMAIN      102    119       CONNECTING PEPTIDE.
FT   DOMAIN      120    161       MOTIF 3.
FT   DOMAIN      162    205       MOTIF 4.
FT   DOMAIN      206    218       C-TERMINAL ARM.
SQ   SEQUENCE   218 AA;  24929 MW;  78C77D24CBB3F8E2 CRC64;


  Query Match             70.8%;  Score 34;  DB 1;  Length 218;
  Best Local Similarity   87.5%;  Pred. No. 15;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 LSGACLNL 9
          ||||| ||
Db     33 LSGACPNL 40


RESULT   6
GSHR_PSEAE
ID   GSHR_PSEAE     STANDARD;      PRT;   451 AA.
AC   P23189;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   GLUTATHIONE REDUCTASE (EC 1.6.4.2) (GR) (GRASE).
GN   GOR.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PAO8;
RX   MEDLINE; 91194546.
RA   Perry A.C.F., Ni Bhriain N., Brown N.L., Rouch D.A.;
RT   "Molecular characterization of the gor gene encoding glutathione
RT   reductase from Pseudomonas aeruginosa: determinants of substrate
RT   specificity among pyridine nucleotide-disulphide oxidoreductases.";
RL   Mol. Microbiol. 5:163-171(1991).
CC   -!- FUNCTION: MAINTAIN HIGH LEVELS OF REDUCED GLUTATHIONE IN THE
CC       CYTOSOL.
CC   -!- CATALYTIC ACTIVITY: NADPH + OXIDIZED GLUTATHIONE = NADP(+) +
CC       2 GLUTATHIONE.
CC   -!- COFACTOR: FAD.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
CC       OXIDOREDUCTASES CLASS-I.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
```

```
DR   EMBL; X54201; CAA38122.1; -.
DR   PIR; S15236; S15236.
DR   HSSP; P00390; 1ALG.
DR   INTERPRO; IPR001100; -.
DR   INTERPRO; IPR001327; -.
DR   PFAM; PF00070; pyr_redox; 1.
DR   PRINTS; PR00368; FADPNR.
DR   PRINTS; PR00411; PNDRDTASEI.
DR   PROSITE; PS00076; PYRIDINE_REDOX_1; 1.
KW   Redox-active center; Oxidoreductase; Flavoprotein; FAD; NADP.
FT   NP_BIND       6     36       FAD (ADP PART) (PROBABLE).
FT   DISULFID     42     47       REDOX-ACTIVE.
FT   NP_BIND     292    302       FAD (FLAVIN PART) (BY SIMILARITY).
FT   ACT_SITE    436    436       PROTON ACCEPTOR/DONOR (POTENTIAL).
SQ   SEQUENCE   451 AA;  49237 MW;  2924599996DB98EA CRC64;


  Query Match             70.8%;  Score 34;  DB 1;  Length 451;
  Best Local Similarity   55.6%;  Pred. No. 28;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              || | |:|:
Db         37 YLGGTCVNV 45


RESULT   7
AQP8_MOUSE
ID   AQP8_MOUSE     STANDARD;      PRT;   261 AA.
AC   P56404;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   AQUAPORIN 8.
GN   AQP8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RX   MEDLINE; 98049830.
RA   Ma T., Yang B., Verkman A.S.;
RT   "Cloning of a novel water and urea-permeable aquaporin from mouse
RT   expressed strongly in colon, placenta, liver, and heart.";
RL   Biochem. Biophys. Res. Commun. 240:324-328(1997).
CC   -!- FUNCTION: FORMS A WATER-SPECIFIC CHANNEL; MERCURY-SENSITIVE. ALSO
CC       PERMEABLE TO UREA BUT NOT TO GLYCEROL.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: COLON, LIVER, HEART AND PLACENTA.
CC   -!- SIMILARITY: BELONGS TO THE TRANSMEMBRANE CHANNEL MIP FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF018952; AAB68847.1; -.
DR   MGD; MGI:1195271; AQP8.
DR   INTERPRO; IPR000425; -.
DR   PFAM; PF00230; MIP; 1.
DR   PRINTS; PR00783; MINTRINSICP.
DR   PROSITE; PS00221; MIP; 1.
KW   Transport; Transmembrane.
FT   DOMAIN        1     36       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     37     57       POTENTIAL.
FT   DOMAIN       58     84       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     85    105       POTENTIAL.
FT   DOMAIN      106    107       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    108    128       POTENTIAL.
```

```
FT   DOMAIN      129    156       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    157    177       POTENTIAL.
FT   DOMAIN      178    183       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    184    204       POTENTIAL.
FT   DOMAIN      205    228       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    229    249       POTENTIAL.
FT   DOMAIN      250    261       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    139    139       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   261 AA;  27797 MW;  9CB48ACA1FB5E589 CRC64;

  Query Match             68.8%;  Score 33;  DB 1;  Length 261;
  Best Local Similarity   71.4%;  Pred. No. 26;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LSGACLN 8
              :||||:|
Db        204 ISGACMN 210


RESULT   8
AQP8_RAT
ID   AQP8_RAT       STANDARD;      PRT;   263 AA.
AC   P56405;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   AQUAPORIN 8.
GN   AQP8.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=TESTIS;
RX   MEDLINE; 97445104.
RA   Ishibashi K., Kuwahara M., Kageyama Y., Tohsaka A., Marumo F.,
RA   Sasaki S.;
RT   "Cloning and functional expression of a second new aquaporin
RT   abundantly expressed in testis.";
RL   Biochem. Biophys. Res. Commun. 237:714-718(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER, AND PANCREAS;
RX   MEDLINE; 98043735.
RA   Koyama Y., Yamamoto T., Kondo D., Funaki H., Yaoita E., Kawasaki K.,
RA   Sato N., Hatakeyama K., Kihara I.;
RT   "Molecular cloning of a new aquaporin from rat pancreas and liver.";
RL   J. Biol. Chem. 272:30329-30333(1997).
CC   -!- FUNCTION: FORMS A WATER-SPECIFIC CHANNEL; MERCURY-SENSITIVE. IT
CC       MAY HAVE AN IMPORTANT ROLE IN SPERMATOGENESIS, IN FERTILIZATION,
CC       AND IN THE SECRETION OF PANCRETIC JUICE AND SALIVA.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SPERM, PANCREAS AND LIVER.
CC       SOME EXPRESSION HAS BEEN FOUND IN SALIVARY GLAND AND ABSORPTIVE
CC       COLONIC EPITHELIAL CELLS.
CC   -!- SIMILARITY: BELONGS TO THE TRANSMEMBRANE CHANNEL MIP FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB005547; BAA21918.1; -.
DR   EMBL; AF007775; AAC53463.1; -.
DR   INTERPRO; IPR000425; -.
DR   PFAM; PF00230; MIP; 1.
DR   PRINTS; PR00783; MINTRINSICP.
DR   PROSITE; PS00221; MIP; 1.
KW   Transport; Transmembrane.
```

```
FT   DOMAIN        1     38       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     39     59       POTENTIAL.
FT   DOMAIN       60     86       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     87    107       POTENTIAL.
FT   DOMAIN      108    109       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    110    130       POTENTIAL.
FT   DOMAIN      131    158       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    159    179       POTENTIAL.
FT   DOMAIN      180    185       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    186    206       POTENTIAL.
FT   DOMAIN      207    230       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    231    251       POTENTIAL.
FT   DOMAIN      252    263       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    141    141       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   263 AA;  28055 MW;  B34AC8ACFA596BD2 CRC64;

  Query Match             68.8%;  Score 33;  DB 1;  Length 263;
  Best Local Similarity   71.4%;  Pred. No. 27;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGACLN 8
           :||||:|
Db     206 ISGACMN 212


RESULT   9
1C28_PANTR
ID   1C28_PANTR     STANDARD;      PRT;   346 AA.
AC   P16215;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN, CH28 ALPHA CHAIN PRECURSOR.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90201944.
RA   Lawlor D.A., Warren E., Ward F.E., Parham P.;
RT   "Comparison of class I MHC alleles in humans and apes.";
RL   Immunol. Rev. 113:147-185(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88319000.
RA   Lawlor D.A., Ward F.E., Ennis P.D., Jackson A.P., Parham P.;
RT   "HLA-A and B polymorphisms predate the divergence of humans and
RT   chimpanzees.";
RL   Nature 335:268-271(1988).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC       THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M30685; AAA87973.1; -.
DR   HSSP; P03989; 1HSA.
DR   INTERPRO; IPR000495; -.
DR   INTERPRO; IPR001039; -.
DR   INTERPRO; IPR003006; -.
DR   PFAM; PF00129; MHC_I; 1.
DR   PFAM; PF00047; ig; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL        1     21
FT   CHAIN        22    346       CHLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT                                CH28 ALPHA CHAIN.
FT   DOMAIN       22    111       EXTRACELLULAR ALPHA-1.
FT   DOMAIN      112    203       EXTRACELLULAR ALPHA-2.
FT   DOMAIN      204    295       EXTRACELLULAR ALPHA-3.
FT   DOMAIN      296    305       CONNECTING PEPTIDE.
FT   TRANSMEM    306    329
FT   DOMAIN      330    346       CYTOPLASMIC TAIL.
FT   DISULFID    122    185       BY SIMILARITY.
FT   DISULFID    224    280       BY SIMILARITY.
FT   CARBOHYD    107    107       N-LINKED (GLCNAC. . .) (BY SIMILARITY).
SQ   SEQUENCE   346 AA;  39084 MW;  F83E882D5C2E0971 CRC64;

  Query Match             68.8%;  Score 33;  DB 1;  Length 346;
  Best Local Similarity   66.7%;  Pred. No. 34;
  Matches    6;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
           || | || |
Db     180 YLEGECLEL 188


RESULT  10
ADH1_PICST
ID   ADH1_PICST     STANDARD;      PRT;   348 AA.
AC   O00097;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   ALCOHOL DEHYDROGENASE I (EC 1.1.1.1) (ADH 2).
GN   ADH1 OR ADH2.
OS   Pichia stipitis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Pichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS 6054;
RX   MEDLINE; 98207839.
RA   Cho J.Y., Jeffries T.W.;
RT   "Pichia stipitis genes for alcohol dehydrogenase with fermentative
RT   and respiratory functions.";
RL   Appl. Environ. Microbiol. 64:1350-1358(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS 5774;
RX   MEDLINE; 99019018.
RA   Passoth V., Schaefer B., Liebel B., Weierstall T., Klinner U.;
RT   "Molecular cloning of alcohol dehydrogenase genes of the yeast Pichia
RT   stipitis and identification of the fermentative ADH.";
RL   Yeast 14:1311-1325(1998).
CC   -!- FUNCTION: CONVERTS ETHANOL TO ACETALDEHYDE AND PLAYS A MAJOR ROLE
CC       IN XYLOSE FERMENTATION.
CC   -!- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.
CC   -!- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY. BINDS TWO ZINC IONS, ONE
CC       ESSENTIAL FOR ACTIVITY, THE OTHER FOR THE STRUCTURAL INTEGRITY.
CC   -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF008245; AAC49991.1; -.
DR   EMBL; Y13397; CAA73827.1; -.
DR   INTERPRO; IPR002085; -.
DR   INTERPRO; IPR002328; -.
```

```
DR   PFAM; PF00107; adh_zinc; 1.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc; NAD; Multigene family.
FT   METAL        44     44       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        67     67       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL        98     98       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       101    101       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       104    104       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       112    112       ZINC (SECOND ATOM) (BY SIMILARITY).
FT   METAL       154    154       ZINC (CATALYTIC) (BY SIMILARITY).
SQ   SEQUENCE   348 AA;  36520 MW;  49C06B545D5350F4 CRC64;


  Query Match             68.8%;  Score 33;  DB 1;  Length 348;
  Best Local Similarity   62.5%;  Pred. No. 34;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGACLN 8
              :|:|:|||
Db         93 WLNGSCLN 100


RESULT  11
HLAF_MACMU
ID   HLAF_MACMU     STANDARD;      PRT;   348 AA.
AC   P33617;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, ALPHA CHAIN F PRECURSOR (HLA F
DE   ANTIGEN) (LEUKOCYTE ANTIGEN F).
GN   HLA-F OR HLAF.
OS   Macaca mulatta (Rhesus macaque).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93246295.
RA   Otting N., Bontrop R.E.;
RT   "Characterization of the rhesus macaque (Macaca mulatta) equivalent
RT   of HLA-F.";
RL   Immunogenetics 38:141-145(1993).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC       THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z21819; CAA79885.1; -.
DR   PIR; S29990; S29990.
DR   HSSP; P03989; 1HSA.
DR   INTERPRO; IPR000495; -.
DR   INTERPRO; IPR001039; -.
DR   INTERPRO; IPR003006; -.
DR   PFAM; PF00129; MHC_I; 1.
DR   PFAM; PF00047; ig; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL        1     21       BY SIMILARITY.
FT   CHAIN        22    348       HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT                                ALPHA CHAIN F.
FT   DOMAIN       22    113       EXTRACELLULAR ALPHA-1.
FT   DOMAIN      114    205       EXTRACELLULAR ALPHA-2.
FT   DOMAIN      206    297       EXTRACELLULAR ALPHA-3.
FT   DOMAIN      298    307       CONNECTING PEPTIDE.
FT   TRANSMEM    308    331
FT   DOMAIN      332    348       CYTOPLASMIC TAIL.
FT   DISULFID    124    187       BY SIMILARITY.
FT   DISULFID    226    282       BY SIMILARITY.
FT   CARBOHYD    109    109       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   348 AA;  39300 MW;  77BD7E3B9B11E0F7 CRC64;

  Query Match             68.8%;  Score 33;  DB 1;  Length 348;
  Best Local Similarity   66.7%;  Pred. No. 34;
  Matches    6;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              || | || |
Db        182 YLEGECLEL 190


RESULT  12
HLAF_HUMAN
ID   HLAF_HUMAN     STANDARD;      PRT;   362 AA.
AC   P30511;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, ALPHA CHAIN F PRECURSOR (HLA F
DE   ANTIGEN) (LEUKOCYTE ANTIGEN F) (CDA12).
GN   HLA-F OR HLAF OR HLA-5.4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90111605.
RA   Geraghty D.E., Wei X., Orr H.T., Koller B.H.;
RT   "Human leukocyte antigen F (HLA-F). An expressed HLA gene composed of
RT   a class I coding sequence linked to a novel transcribed repetitive
RT   element.";
RL   J. Exp. Med. 171:1-18(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91197889.
RA   Lury D., Epstein H., Holmes N.;
RT   "The human class I MHC gene HLA-F is expressed in lymphocytes.";
RL   Int. Immunol. 2:531-537(1990).
CC   -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC       THE IMMUNE SYSTEM.
CC   -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC       MICROGLOBULIN).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X17093; CAA34947.1; -.
DR   PIR; A60384; A60384.
DR   PIR; JL0147; JL0147.
DR   HSSP; P03989; 1HSA.
DR   MIM; 143110; -.
DR   INTERPRO; IPR000495; -.
DR   INTERPRO; IPR001039; -.
DR   INTERPRO; IPR003006; -.
DR   PFAM; PF00129; MHC_I; 1.
DR   PFAM; PF00047; ig; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   MHC I; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL        1     21
FT   CHAIN        22    362       HLA CLASS I HISTOCOMPATIBILITY ANTIGEN,
FT                                ALPHA CHAIN F.
FT   DOMAIN       22    111       EXTRACELLULAR ALPHA-1.
```

```
FT   DOMAIN      112    203       EXTRACELLULAR ALPHA-2.
FT   DOMAIN      204    295       EXTRACELLULAR ALPHA-3.
FT   DOMAIN      296    305       CONNECTING PEPTIDE.
FT   TRANSMEM    306    329
FT   DOMAIN      330    362       CYTOPLASMIC TAIL.
FT   DISULFID    122    185       BY SIMILARITY.
FT   DISULFID    224    280       BY SIMILARITY.
FT   CARBOHYD    107    107       N-LINKED (GLCNAC. . .) (BY SIMILARITY).
SQ   SEQUENCE   362 AA;  40568 MW;  B8DD041F820A34E1 CRC64;


  Query Match             68.8%;  Score 33;  DB 1;  Length 362;
  Best Local Similarity   66.7%;  Pred. No. 35;
  Matches    6;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
           || | || |
Db     180 YLEGECLEL 188


RESULT  13
Y093_HAEIN
ID   Y093_HAEIN     STANDARD;      PRT;   368 AA.
AC   P44509;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL PROTEIN HI0093.
GN   HI0093.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- SIMILARITY: BELONGS TO THE CDAR FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U32695; AAC21771.1; -.
DR   TIGR; HI0093; -.
KW   Hypothetical protein.
SQ   SEQUENCE   368 AA;  42251 MW;  C77F1C9EF043B89A CRC64;


  Query Match             68.8%;  Score 33;  DB 1;  Length 368;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GACLNL 9
           ||||||
Db     235 GACLNL 240
```

```
RESULT  14
DLDH_MYCGE
ID   DLDH_MYCGE     STANDARD;      PRT;   457 AA.
AC   P47513; Q49233;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   DIHYDROLIPOAMIDE DEHYDROGENASE (EC 1.8.1.4) (E3 COMPONENT OF PYRUVATE
DE   COMPLEX).
GN   PDHD OR MG271.
OS   Mycoplasma genitalium.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE; 96026346.
RA   Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA   Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA   Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA   Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA   Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA   Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT   "The minimal gene complement of Mycoplasma genitalium.";
RL   Science 270:397-403(1995).
RN   [2]
RP   SEQUENCE OF 306-404 FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE; 94075230.
RA   Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT   "A survey of the Mycoplasma genitalium genome by using random
RT   sequencing.";
RL   J. Bacteriol. 175:7918-7930(1993).
CC   -!- FUNCTION: LIPOAMIDE DEHYDROGENASE IS A COMPONENT OF THE ALPHA-
CC       KETOACID DEHYDROGENASE COMPLEXES (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: DIHYDROLIPOAMIDE + NAD(+) = LIPOAMIDE + NADH.
CC   -!- COFACTOR: FAD (BY SIMILARITY).
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
CC       OXIDOREDUCTASES CLASS-I.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U39706; AAC71493.1; -.
DR   EMBL; U01784; AAD10605.1; -.
DR   HSSP; P11959; 1EBD.
DR   TIGR; MG271; -.
DR   INTERPRO; IPR001100; -.
DR   INTERPRO; IPR001327; -.
DR   PFAM; PF00070; pyr_redox; 1.
DR   PRINTS; PR00368; FADPNR.
DR   PRINTS; PR00411; PNDRDTASEI.
DR   PROSITE; PS00076; PYRIDINE_REDOX_1; 1.
KW   Redox-active center; Glycolysis; Oxidoreductase; NAD; Flavoprotein;
KW   FAD.
FT   NP_BIND       4     34       FAD (ADP PART) (PROBABLE).
FT   DISULFID     40     45       REDOX-ACTIVE (BY SIMILARITY).
FT   NP_BIND     293    303       FAD (FLAVIN PART) (BY SIMILARITY).
FT   ACT_SITE    437    437       BY SIMILARITY.
FT   CONFLICT    388    392       FVKMM -> LSRWC (IN REF. 2).
SQ   SEQUENCE   457 AA;  50119 MW;  2A0B3F11370AD072 CRC64;


  Query Match             68.8%;  Score 33;  DB 1;  Length 457;
  Best Local Similarity   55.6%;  Pred. No. 43;
```

```
  Matches       5;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              |  | |||:
Db         35 YFGGVCLNV 43


RESULT  15
DLDH_MYCPN
ID   DLDH_MYCPN     STANDARD;      PRT;   457 AA.
AC   P75393;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   DIHYDROLIPOAMIDE DEHYDROGENASE (EC 1.8.1.4) (E3 COMPONENT OF PYRUVATE
DE   COMPLEX).
GN   PDHD.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE; 97105885.
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -!- FUNCTION: LIPOAMIDE DEHYDROGENASE IS A COMPONENT OF THE ALPHA-
CC       KETOACID DEHYDROGENASE COMPLEXES (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: DIHYDROLIPOAMIDE + NAD(+) = LIPOAMIDE + NADH.
CC   -!- COFACTOR: FAD (BY SIMILARITY).
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
CC       OXIDOREDUCTASES CLASS-I.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE000044; AAB96096.1; -.
DR   HSSP; P11959; 1EBD.
DR   INTERPRO; IPR001100; -.
DR   PFAM; PF00070; pyr_redox; 1.
DR   PRINTS; PR00411; PNDRDTASEI.
DR   PROSITE; PS00076; PYRIDINE_REDOX_1; 1.
KW   Redox-active center; Glycolysis; Oxidoreductase; NAD; Flavoprotein;
KW   FAD.
FT   NP_BIND       4     34       FAD (ADP PART) (PROBABLE).
FT   DISULFID     40     45       REDOX-ACTIVE (BY SIMILARITY).
FT   NP_BIND     293    303       FAD (FLAVIN PART) (BY SIMILARITY).
FT   ACT_SITE    437    437       BY SIMILARITY.
SQ   SEQUENCE   457 AA;  49437 MW;  EB044FD676F3F28E CRC64;


  Query Match             68.8%;  Score 33;  DB 1;  Length 457;
  Best Local Similarity   55.6%;  Pred. No. 43;
  Matches       5;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
              |  | |||:
Db         35 YFGGVCLNV 43


Search completed: December 16, 2000, 03:05:21
```

```
Job time: 8041 sec
```

```
                        GenCore version 4.5
              Copyright (c) 1993 - 2000  Compugen Ltd.


OM protein - protein search, using sw model

Run on:         December 16, 2000, 00:51:14 ; Search time 108.84 Seconds
                                               (without alignments)
                                               2.827 Million cell updates/sec

Title:          US-09-529-121-5
Perfect score:  48
Sequence:       1 YLSGACLNL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters:      268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_36:*
                1:  /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
                2:  /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
                3:  /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
                4:  /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
                5:  /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
                6:  /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
                7:  /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
                8:  /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
                9:  /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
                10:  /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
                11:  /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
                12:  /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
                13:  /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
                14:  /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
                15:  /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
                16:  /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
                17:  /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
                18:  /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
                19:  /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
                20:  /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
                21:  /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                 %
Result          Query
  No.   Score   Match Length DB  ID                    Description
 --------------------------------------------------------------------------
   1      48   100.0      9  20  Y09529                 Carcinoembryonic a
   2      36    75.0      9  18  W39723                 Human carcina-embr
   3      36    75.0      9  19  W77134                 CEA synthetic pept
   4      36    75.0      9  19  W70045                 CEA derived HLA-A2
   5      36    75.0      9  20  Y47655                 Immunogenic peptid
   6      36    75.0      9  20  Y09525                 Carcinoembryonic a
   7      36    75.0      9  20  Y09526                 Carcinoembryonic a
   8      36    75.0     10  20  Y46555                 Immunogenic peptid
   9      36    75.0    107  20  W86133                 Protein sequence o
  10      36    75.0    178  10  P93499                 Sequence of carcin
  11      36    75.0    468  16  R77436                 BGP (1-314)/CEA (4
  12      36    75.0    493  16  R77435                 BGP (1-314)/CEA (4
  13      36    75.0    509  16  R77437                 BGP (1-314)/CEA (4
  14      36    75.0    511  16  R77438                 BGP (1-314)/CEA (4
  15      36    75.0    642  15  R60619                 Carcinoembryonic a
  16      36    75.0    663  17  R98519                 Immunogenic carcin
  17      36    75.0    698  9  P81229                 Carcinoembryonic a
  18      36    75.0    698  16  R65168                 Carcinoembryonic a
  19      36    75.0    698  18  W22844                 Human carcinoembry
  20      36    75.0    702  9  P81222                 Carcinoembryonic a
  21      36    75.0    702  10  P94014                 Carcinoembryonic c
  22      36    75.0    702  10  P93999                 Amino acid sequenc
  23      36    75.0    702  15  R54713                 Carcinoembryonic a
  24      36    75.0    702  17  W06872                 Carcinoembryonic a
  25      36    75.0    702  20  W83137                 CEA protein.  Homo
  26      36    75.0    734  17  W00182                 Carcinoembryonic a
  27      36    75.0   3329  21  Y56029                 Human BRCA2 tumour
  28      35    72.9    210  19  W39177                 Human PKD1 protein
  29      35    72.9   4302  17  W00870                 Polycystic kidney
  30      35    72.9   4302  19  W33396                 Human PKD1 polypep
  31      35    72.9   4302  19  W23830                 Human PKD1 protein
  32      35    72.9   4303  17  R90302                 Polycystic kidney
  33      35    72.9   4339  16  R75916                 Polycystic kidney
  34      35    72.9   4339  17  R87539                 Polycystic kidney
  35      34    70.8      9  20  Y09527                 Carcinoembryonic a
  36      34    70.8      9  20  Y09528                 Carcinoembryonic a
  37      34    70.8     79  18  W28107                 Staphylococcus aur
  38      33    68.8      9  21  Y54173                 HLA binding peptid
  39      33    68.8    183  20  Y27573                 Human secreted pro
  40      33    68.8    648  18  W24472                 Acetolactate synth
  41      33    68.8    648  18  W24473                 Herbicide-sensitiv
  42      33    68.8    660  13  R28615                 Amino acids 569-57
  43      33    68.8    662  13  R28611                 Amino acids 201-20
  44      33    68.8    664  13  R28613                 Amino acids 348-35
  45      33    68.8    665  13  R28614                 Amino acids 373-37


                                 ALIGNMENTS


RESULT   1
Y09529
ID   Y09529 standard; peptide; 9 AA.
XX
AC   Y09529;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:5.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
```

```
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;


  Query Match             100.0%;  Score 48;  DB 20;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGACLNL 9
            |||||||||
Db        1 ylsgaclnl 9


RESULT   2
W39723
ID   W39723 standard; peptide; 9 AA.
XX
AC   W39723;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Human carcina-embryonic antigen (CEA) peptide (pos. 571-579).
XX
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
XX
OS   Homo sapiens.
XX
PN   WO9741440-A1.
XX
PD   06-NOV-1997.
XX
PF   28-APR-1997;   97WO-NL00229.
XX
PR   23-DEC-1996;   96EP-0203670.
PR   26-APR-1996;   96EP-0201145.
XX
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
PI   Kast WM,  Melief CJM,  Offringa R,  Toes REM,  Van Der Burg SH;
XX
DR   WPI; 1997-549891/50.
XX
PT   Method of selecting T cell peptide epitope(s) - by measuring the
PT   stability of HLA class I-peptide complexes on intact B cells
XX
PS   Example 3; Page 85; 109pp; English.
XX
CC   Peptides W39430-W39734 are used in a novel method for the selection of
CC   immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC   method involves the identification of peptide sequences capable of
CC   binding to an HLA (human leukocyte antigen) class I molecule and
CC   measuring the binding of this epitope peptide to the HLA class I peptide.
CC   The stability of binding of the peptide and MHC (major histocompatablilty
CC   complex) class I molecule is measured on intact human B cells carrying
CC   the MHC molecule at their cell surfaces. The method can be used to select
CC   peptide epitopes for generating vaccines against a disease associated
CC   with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC   especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC   immune responses. Peptide W39723 is derived from the human
CC   carcino-embryonic antigen (CEA) and has the ability to bind to the human
CC   MHC Class I allele HLA-A2.1.
XX
SQ   Sequence   9 AA;


  Query Match             75.0%;  Score 36;  DB 18;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGACLNL 9
              ||||| |||
Db        1 ylsganlnl 9


RESULT   3
W77134
ID   W77134 standard; peptide; 9 AA.
XX
AC   W77134;
XX
DT   16-NOV-1998  (first entry)
XX
DE   CEA synthetic peptide epitope 1.
XX
KW   Tyrosinase; tyrosinase cytotoxic lymphocyte response;
KW   cytotoxic T lymphocyte; cysteine-depleted; melanoma.
XX
OS   Synthetic.
XX
PN   WO9833810-A2.
XX
PD   06-AUG-1998.
XX
PF   29-JAN-1998;   98WO-US01592.
XX
PR   30-JAN-1997;   97US-0037781.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI   Engelhard VH,  Hunt DF,  Kittlesen D,  Slingluff CL;
XX
DR   WPI; 1998-437388/37.
XX
PT   Disease specific immunogen  - comprises disease specific cytotoxic T
PT   lymphocyte epitope used to elicit melanoma specific CTL response
XX
PS   Disclosure; Page 27; 93pp; English.
XX
CC   The peptide epitope W77119-W77138 were created for human tumour-specific
CC   cytotoxic T lymphocyte response.  These peptides are are cysteine-
CC   depleted mutants of a native disease-specific CTL epitope.  The cysteine-
CC   depleted CTL epitopes elicit a stronger or more specific CTL response
CC   than the native epitope.  The epitopes can be used in a disease-specific
CC   immunogen to protect a mammal against disease in particular melanomas.
CC   The peptides may also be used to screen a sample for the presence of
CC   an antigen with the same epitope, or with a different cross-reactive
CC   epitope.
XX
SQ   Sequence   9 AA;


  Query Match             75.0%;  Score 36;  DB 19;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGACLNL 9
              ||||| |||
Db        1 ylsganlnl 9
```

```
RESULT   4
W70045
ID   W70045 standard; peptide; 9 AA.
XX
AC   W70045;
XX
DT   22-OCT-1998  (first entry)
XX
DE   CEA derived HLA-A2.1 binding peptide 2 (residues 605-613).
XX
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW   antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW   viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW   fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9833888-A1.
XX
PD   06-AUG-1998.
XX
PF   30-JAN-1998;   98WO-US01959.
XX
PR   31-JAN-1997;   97US-0036696.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Celis E,  Sette A,  Sidney J,  Southwood S,  Tsai V;
XX
DR   WPI; 1998-437445/37.
XX
PT   Production of antigen-specific cytotoxic T cells - by incubating
PT   immunogenic peptide(s) from antigen that binds class I major
PT   histocompatibility complex molecules with pre-treated antigen
PT   presenting cells
XX
PS   Example 6; Page 75; 104pp; English.
XX
CC   Sequences shown in W70044 to W70052 represent peptides derived from
CC   carcinoembryonic antigen (CEA). The peptides can bind to a human
CC   leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method
CC   of invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC   vitro. The method comprises contacting immunogenic peptides from an
CC   antigen that binds class I major histocompatibility complex (MHC)
CC   molecules with antigen presenting cells (APCs) pretreated with
CC   pretreatment growth factors, and incubating the APCs with purified CD8
CC   cells in the presence of at least 2 incubation growth factors, thereby
CC   producing antigen-specific CTLs. A method for specifically killing
CC   target cells in a human patient is also provided which comprises
CC   obtaining a fluid sample containing CTLs from a patient, contacting the
CC   cytotoxic T cells with APCs pretreated with pre-treatment growth factors,
CC   where the APCs comprise class I MHC molecules. The pretreated APCs are
CC   incubated with the cytotoxic growth factors, thereby producing activated
CC   CTLs which are contacted with a carrier to form a composition. The
CC   composition can then be administered to the patient. The activated CTLs
CC   can be used for treating cancers, immune disorders, viral infections,
CC   AIDS, hepatitis, bacterial infection, fungal infection, malaria or
CC   tuberculosis.
XX
SQ   Sequence   9 AA;

  Query Match             75.0%;  Score 36;  DB 19;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGACLNL 9
            ||||| |||
Db        1 ylsganlnl 9
```

```
RESULT   5
Y47655
ID   Y47655 standard; Peptide; 9 AA.
XX
AC   Y47655;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2266.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 118; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;

  Query Match             75.0%;  Score 36;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGACLNL 9
            ||||| |||
Db        1 ylsganlnl 9
```

```
RESULT   6
Y09525
ID   Y09525 standard; peptide; 9 AA.
XX
AC   Y09525;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist CAP-1.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 1; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             75.0%;  Score 36;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
           ||||| |||
Db       1 ylsganlnl 9

RESULT   7
Y09526
ID   Y09526 standard; peptide; 9 AA.
XX
AC   Y09526;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:2.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             75.0%;  Score 36;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
           ||||| |||
Db       1 ylsgadlnl 9

RESULT   8
Y46555
ID   Y46555 standard; Peptide; 10 AA.
XX
AC   Y46555;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #1166.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
```

```
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 76; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   10 AA;

  Query Match             75.0%;  Score 36;  DB 20;  Length 10;
  Best Local Similarity   88.9%;  Pred. No. 0.51;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGACLNL 9
          ||||| |||
Db      2 ylsganlnl 10


RESULT   9
W86133
ID   W86133 standard; Protein; 107 AA.
XX
AC   W86133;
XX
DT   03-MAR-1999  (first entry)
XX
DE   Protein sequence of vaccine 2 708 Vl.
XX
KW   Non-immunogenic; epitope; T-cell; immunogenicity; immune system; SK;
KW   immunoglobulin; therapeutic; streptokinase; vaccine; 708.
XX
OS   Homo sapiens.
XX
PN   WO9852976-A1.
XX
PD   26-NOV-1998.
XX
PF   21-MAY-1998;   98WO-GB01473.
XX
PR   14-APR-1998;   98GB-0007751.
PR   21-MAY-1997;   97GB-0010480.
```

```
PR   31-JUL-1997;   97GB-0016197.
PR   28-NOV-1997;   97GB-0025270.
PR   02-DEC-1997;   97US-0067235.
XX
PA   (BIOV-) BIOVATION LTD.
XX
PI   Carr FJ;
XX
DR   WPI; 1999-045301/04.
XX
PT   Reducing immunogenicity of proteins - by modifying the amino acid
PT   sequence of the protein to eliminate potential epitopes for T-cells
PT   of a given species
XX
PS   Example 4; Fig 19; 77pp; English.
XX
CC   The invention relates to a method for the production of non-immunogenic
CC   proteins. The method comprises determining at least part of the amino
CC   acid sequence of the protein; (b) identifying in the amino acid sequence
CC   one or more potential epitopes for T-cells (T-cell epitopes) of the given
CC   species; and (c) modifying the amino acid sequence to eliminate at least
CC   one of the T-cell epitopes identified in step (b) thereby to eliminate or
CC   reduce the immunogenicity of the protein when exposed to the immune
CC   system of the given species. A method of analysing a pre-existing protein
CC   to predict the basis for immunogenic responses is also provided. The
CC   methods can be used particularly for reducing the immunogenicity of
CC   immunoglobulins or therapeutic proteins, e.g. Streptokinase (SK). The
CC   products can be used for diagnosis and therapy. The present sequence
CC   represents the protein sequence of vaccine 2 708 Vl.
XX
SQ   Sequence   107 AA;

  Query Match             75.0%;  Score 36;  DB 20;  Length 107;
  Best Local Similarity   88.9%;  Pred. No. 6;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGACLNL 9
          ||||| |||
Db     89 ylsganlnl 97


RESULT  10
P93499
ID   P93499 standard; protein; 178 AA.
XX
AC   P93499;
XX
DT   08-MAY-1990  (first entry)
XX
DE   Sequence of carcinoembryonic antigen domain III.
XX
KW   Carcinoembryonic antigen; domain III; domain A; domain B.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..89
FT                   /note="domain A"
FT   Domain          90..178
FT                   /note="domain B"
XX
PN   EP343946-A.
XX
PD   29-NOV-1989.
XX
PF   24-MAY-1989;   89EP-0305232.
XX
PR   25-MAY-1988;   88US-0198289.
XX
PA   (CITY ) CITY OF HOPE.
XX
PI   Shively JE;
XX
DR   WPI; 1989-349991/48.
```

```
DR   N-PSDB; N92449.
XX
PT   Carcinoembryonic antigen fragments - used in assays to determine the
PT   presence and amt. of the antigen in samples also contg. related antigens.
XX
PS   Disclosure; page 4; 15pp; English.
XX
CC   CEA fragments can be used in assays to determine the presence and amt. of
CC   CEA in samples which also may contain related antigens including its
CC   normal cross-reacting antigen or the 128 kD antigen.
XX
SQ   Sequence   178 AA;


  Query Match             75.0%;  Score 36;  DB 10;  Length 178;
  Best Local Similarity   88.9%;  Pred. No. 10;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
             ||||| |||
Db     107 ylsganlnl 115


RESULT  11
R77436
ID   R77436 standard; Protein; 468 AA.
XX
AC   R77436;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-643) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..468
FT                   /note= "CEA (490-643)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 16; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
```

```
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   468 AA;

  Query Match             75.0%;  Score 36;  DB 16;  Length 468;
  Best Local Similarity   88.9%;  Pred. No. 28;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
             ||||| |||
Db     396 ylsganlnl 404


RESULT  12
R77435
ID   R77435 standard; Protein; 493 AA.
XX
AC   R77435;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-C-terminal) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..493
FT                   /note= "CEA (490-C-terminal)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 15; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
```

```
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   493 AA;

  Query Match             75.0%;  Score 36;  DB 16;  Length 493;
  Best Local Similarity   88.9%;  Pred. No. 29;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGACLNL 9
              ||||| |||
Db      396 ylsganlnl 404

RESULT  13
R77437
ID   R77437 standard; Protein; 509 AA.
XX
AC   R77437;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-644)/BGP (391-430) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..469
FT                   /note= "CEA (490-644)"
FT   Protein         470..509
FT                   /note= "BGP (391-430)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 17; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
```

```
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   509 AA;

  Query Match             75.0%;  Score 36;  DB 16;  Length 509;
  Best Local Similarity   88.9%;  Pred. No. 30;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGACLNL 9
              ||||| |||
Db      396 ylsganlnl 404

RESULT  14
R77438
ID   R77438 standard; Protein; 511 AA.
XX
AC   R77438;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-642)/BGP (387-430) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..467
FT                   /note= "CEA (490-642)"
FT   Protein         468..511
FT                   /note= "BGP (387-430)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 18; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
```

```
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   511 AA;


  Query Match             75.0%;  Score 36;  DB 16;  Length 511;
  Best Local Similarity   88.9%;  Pred. No. 30;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
             ||||| |||
Db     396 ylsganlnl 404


RESULT  15
R60619
ID   R60619 standard; Protein; 642 AA.
XX
AC   R60619;
XX
DT   10-MAY-1995  (first entry)
XX
DE   Carcinoembryonic antigen glycoprotein.
XX
KW   Carcinoembryonic antigen; CEA; neoplastic diseases.
XX
OS   Homo sapiens.
XX
PN   EP618292-A.
XX
PD   05-OCT-1994.
XX
PF   15-MAR-1994;   94EP-0103986.
XX
PR   25-MAR-1993;   93EP-0810214.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Mach J,  Pelegrin A,  Terskikh A;
XX
DR   WPI; 1994-304461/38.
DR   N-PSDB; Q71567.
XX
PT   Carcinoembryonic antigen (CEA) derivs - useful as reagents in
PT   immunoassay for diagnosis of neoplastic diseases
XX
PS   Claim 2; Page 15; 30pp; English.
XX
CC   Q71567 is the DNA sequence which encodes carcinoembryonic antigen (CEA)
CC   R60619. CEA is free from cross-reactive CEA-like antigens, it is
CC   antigenically indistinguishable from the solution form of CEA shed from
CC   tumour cells, and it is devoid of ethanolamine. R60619 can be used in a
CC   reagent composition for detecting neoplastic diseases in biological
CC   samples, or in an immunoassay process where it can specifically detect
CC   the presence of tumour cells in a biological sample e.g. blood.
XX
SQ   Sequence   642 AA;


  Query Match             75.0%;  Score 36;  DB 15;  Length 642;
  Best Local Similarity   88.9%;  Pred. No. 39;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
             ||||| |||
Db     571 ylsganlnl 579
```

```
Search completed: December 16, 2000, 00:51:15
Job time: 18775 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 04:22:13 ; Search time 111.26 Seconds
                                              (without alignments)
                                              7.553 Million cell updates/sec

Title:          US-09-529-121-5
Perfect score:  48
Sequence:       1 YLSGACLNL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters:      436

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :      SPTREMBL_14:*
               1:  sp_archea:*
               2:  sp_bacteria:*
               3:  sp_fungi:*
               4:  sp_human:*
               5:  sp_invertebrate:*
               6:  sp_mammal:*
               7:  sp_mhc:*
               8:  sp_organelle:*
               9:  sp_phage:*
               10:  sp_plant:*
               11:  sp_rodent:*
               12:  sp_virus:*
               13:  sp_vertebrate:*
               14:  sp_unclassified:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                 %
Result         Query
  No.   Score  Match Length DB  ID                    Description
 -----------------------------------------------------------------------
    1     19    39.6      9  4  Q15999                Q15999 homo sapien
    2     17    35.4      7 12   Q66113                Q66113 cherry leaf
    3     17    35.4      8 13  Q9PS69                Q9ps69 gallus gall
    4     15    31.2      8  3  P87225                P87225 saccharomyc
    5     15    31.2      8  4  Q9UL56                Q9ul56 homo sapien
    6     15    31.2      8  6  Q9TRY3                Q9try3 sus sp. ins
    7     14    29.2      7 13   O42564                O42564 fugu rubrip
    8     14    29.2      8  2  Q9X3K1                Q9x3k1 prochloroco
    9     14    29.2      8 13   P82079                P82079 limnodynast
   10     14    29.2      9  6  Q9TRS0                Q9trs0 oryctolagus
   11     14    29.2      9 11   O35953                O35953 mus musculu
   12     13    27.1      6  4  Q08720                Q08720 homo sapien
   13     13    27.1      8  4  Q9Y4X6                Q9y4x6 homo sapien
   14     13    27.1      8  6  O02831                O02831 oryctolagus
   15     13    27.1      9  2  Q47063                Q47063 escherichia
   16     13    27.1      9  4  Q99887                Q99887 homo sapien
   17     13    27.1      9 11   Q9QZA8                Q9qza8 mus musculu
   18     13    27.1      9 12   Q69473                Q69473 human herpe
   19     12    25.0      8  2  P77556                P77556 escherichia
   20     12    25.0      8  4  Q15901                Q15901 homo sapien
   21     12    25.0      8  4  Q9UMH9                Q9umh9 homo sapien
   22     12    25.0      8  7  Q95213                Q95213 oryctolagus
   23     12    25.0      8 13   Q90498                Q90498 erythrura g
   24     12    25.0      8 13   Q91098                Q91098 manorina me
   25     12    25.0      8 13   P82082                P82082 limnodynast
   26     12    25.0      8 13   P82083                P82083 limnodynast
   27     12    25.0      9  2  Q9R635                Q9r635 chlamydia t
   28     12    25.0      9  4  Q9UCN5                Q9ucn5 homo sapien
   29     12    25.0      9  5  Q27396                Q27396 babesia bov
   30     12    25.0      9  6  Q28112                Q28112 bos taurus
   31     12    25.0      9  6  Q9TRW2                Q9trw2 oryctolagus
   32     12    25.0      9  8  Q31653                Q31653 anser caeru
   33     11    22.9      7 12   Q67113                Q67113 influenza a
   34     11    22.9      7 12   Q9YQ10                Q9yq10 porcine tra
   35     11    22.9      8  2  O32560                O32560 escherichia
   36     11    22.9      8  2  Q9RQ57                Q9rq57 buchnera ap
   37     11    22.9      8  2  Q9RQ49                Q9rq49 buchnera ap
   38     11    22.9      8  4  Q15889                Q15889 homo sapien
   39     11    22.9      8  4  Q9UCN4                Q9ucn4 homo sapien
   40     11    22.9      8  6  Q9XSY1                Q9xsy1 canis famil
   41     11    22.9      8 11   Q9QVF4                Q9qvf4 rattus sp.
   42     11    22.9      8 12   Q83332                Q83332 murine hepa
   43     11    22.9      9  2  Q44001                Q44001 aeromonas e
   44     11    22.9      9  2  Q44377                Q44377 aeromonas t
   45     11    22.9      9  2  Q44468                Q44468 aeromonas v
   46     11    22.9      9  2  Q43928                Q43928 aeromonas c
   47     11    22.9      9  2  Q9R7H9                Q9r7h9 haemophilus
   48     11    22.9      9  2  Q9R5M1                Q9r5m1 staphylococ
   49     11    22.9      9  4  P78484                P78484 homo sapien
   50     11    22.9      9 13   Q92009                Q92009 gallus gall
   51     10    20.8      7  5  Q9VYN9                Q9vyn9 drosophila
   52     10    20.8      7  8  O98866                O98866 spinacia ol
   53     10    20.8      8  2  P72221                P72221 pseudomonas
   54     10    20.8      8  2  Q9R7T2                Q9r7t2 escherichia
   55     10    20.8      8  2  Q9R4M3                Q9r4m3 enterococcu
   56     10    20.8      8  4  Q15890                Q15890 homo sapien
   57     10    20.8      8  4  Q16428                Q16428 homo sapien
   58     10    20.8      8  4  Q9Y4J4                Q9y4j4 homo sapien
   59     10    20.8      8  4  Q9Y4J3                Q9y4j3 homo sapien
   60     10    20.8      8  4  Q9UDZ4                Q9udz4 homo sapien
   61     10    20.8      8  5  Q94623                Q94623 manduca sex
   62     10    20.8      8  8  Q9XNP8                Q9xnp8 boophilus m
   63     10    20.8      8 10   Q40659                Q40659 oryza sativ
   64     10    20.8      8 11   O35835                O35835 rattus norv
   65     10    20.8      8 12   Q66807                Q66807 echovirus 2
   66     10    20.8      8 12   Q83349                Q83349 murine hepa
   67     10    20.8      9  2  Q57328                Q57328 aeromonas s
   68     10    20.8      9  2  Q9R9C4                Q9r9c4 borrelia bu
   69     10    20.8      9  2  Q9R7E8                Q9r7e8 escherichia
   70     10    20.8      9  4  Q16386                Q16386 homo sapien
   71     10    20.8      9  4  Q16276                Q16276 homo sapien
   72     10    20.8      9  4  Q9UQW0                Q9uqw0 homo sapien
   73     10    20.8      9  5  P82003                P82003 bombyx mori
   74     10    20.8      9  5  Q9VV82                Q9vv82 drosophila
   75     10    20.8      9  5  Q9TWD6                Q9twd6 leptinotars

                                 ALIGNMENTS

RESULT   1
Q15999
ID   Q15999 .    PRELIMINARY;      PRT;     9 AA.
AC   Q15999;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   C-KIT PROTOONCOGENE (FRAGMENT).
GN   KIT.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92020918.
RA   Giebel L.B., Spritz R.A.;
RT   "Mutation of the KIT (mast/stem cell growth factor receptor)
RT   protooncogene in human piebaldism.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:8696-8699(1991).
DR   EMBL; S58152; AAB19972.1; -.
DR   EMBL; S58145; AAB19972.1; JOINED.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  875 MW;  D32C74087041AEBD CRC64;

  Query Match             39.6%;  Score 19;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GAC 6
           |||
Db       1 GAC 3

RESULT   2
Q66113
ID   Q66113      PRELIMINARY;      PRT;     7 AA.
AC   Q66113;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   C-TERMINUS OF THE VIRAL REPLICASE (FRAGMENT).
OS   cherry leaf roll virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
OC   Nepovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WALNUT;
RA   Borja M.;
RL   Thesis (1992), Biologia Molecular y Virologia Vegetal, CIT-INIA.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WALNUT;
RX   MEDLINE; 96124520.
RA   Borja M., Sanchez F., Rowhani A., Bruening G., Ponz F.;
RT   "Long, nearly identical untranslated sequences at the 3' terminal
RT   regions of the genomic RNAs of cherry leafroll virus (walnut
RT   strain).";
RL   Virus Genes 10:245-252(1995).
DR   EMBL; Z34265; CAA84019.1; -.
KW   Repeat.
FT   NON_TER       1      1
SQ   SEQUENCE   7 AA;  800 MW;  7417672EBDC6D740 CRC64;

  Query Match             35.4%;  Score 17;  DB 12;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ACL 7
           |||
Db       3 ACL 5

RESULT   3
Q9PS69
ID   Q9PS69      PRELIMINARY;      PRT;     8 AA.
AC   Q9PS69;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   LOW DENSITY LIPOPROTEIN RECEPTOR-RELATED PROTEIN.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
```

```
OC   Gallus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92011685.
RA   Stifani S., Barber D.L., Aebersold R., Steyrer E., Shen X., Nimpf J.,
RA   Schneider W.J.;
RT   "The laying hen expresses two different low density lipoprotein
RT   receptor-related proteins.";
RL   J. Biol. Chem. 266:19079-19087(1991).
SQ   SEQUENCE   8 AA;  846 MW;  C007272DD865BAAA CRC64;

  Query Match             35.4%;  Score 17;  DB 13;  Length 8;
  Best Local Similarity   80.0%;  Pred. No. 3e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGACL 7
           ||| |
Db       3 SGALL 7

RESULT   4
P87225
ID   P87225      PRELIMINARY;      PRT;     8 AA.
AC   P87225;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GIN11 PROTEIN (FRAGMENT).
GN   GIN11.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wedler H., Wedler E., Scharfe M., Wambutt R.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z73169; CAA97518.2; -.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  1019 MW;  4E21A9C449D5B73B CRC64;

  Query Match             31.2%;  Score 15;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLS 3
           |||
Db       1 YLS 3

RESULT   5
Q9UL56
ID   Q9UL56      PRELIMINARY;      PRT;     8 AA.
AC   Q9UL56;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MUTANT NADH-CYTOCHROME B5 REDUCTASE (EC 1.6.2.2) (FRAGMENT).
GN   DIA1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fukumaki Y., Higasa K.;
RT   "Two novel mutations in Thai patients with hereditary
RT   methemoglobinemia types I and II: a subtle amino acid change causes
RT   instability of NADH-cytochrome b5 reductase.";
```

```
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF061830; AAF06818.1; -.
KW   Oxidoreductase.
FT   NON_TER       1      1
FT   VARIANT       9      9       -> R.
SQ   SEQUENCE   8 AA;  888 MW;  76C5B73B5051F6D8 CRC64;


  Query Match             31.2%;  Score 15;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLS 3
            |||
Db        5 YLS 7


RESULT   6
Q9TRY3
ID   Q9TRY3      PRELIMINARY;      PRT;     8 AA.
AC   Q9TRY3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   INSULIN-LIKE GROWTH FACTOR-BINDING PROTEIN-6, IGFBP-6.
OS   Sus sp.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92049376.
RA   Shimasaki S., Gao L., Shimonaka M., Ling N.;
RT   "Isolation and molecular cloning of insulin-like growth factor-binding
RT   protein-6.";
RL   Mol. Endocrinol. 5:938-948(1991).
SQ   SEQUENCE   8 AA;  850 MW;  9FB2CEA37EA7687D CRC64;


  Query Match             31.2%;  Score 15;  DB 6;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        3 SGAC 6
            :| |
Db        1 AGPC 4


RESULT   7
O42564
ID   O42564      PRELIMINARY;      PRT;     7 AA.
AC   O42564;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   TRUNCATED VOLTAGE-GATED SODIUM CHANNEL ALPHA SUBUNIT (FRAGMENT).
GN   SCN8A.
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
OC   Percomorpha; Tetraodontiformes; Tetraodontoidei; Tetraodontidae;
OC   Takifugu.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97442476.
RA   Plummer N.W., McBurney M.W., Meisler M.H.;
RT   "Alternative splicing of the sodium channel SCN8A predicts a truncated
RT   two-domain protein in fetal brain and non-neuronal cells.";
RL   J. Biol. Chem. 272:24008-24015(1997).
DR   EMBL; U97673; AAB80916.1; -.
KW   Ionic channel.
FT   NON_TER       1      1
SQ   SEQUENCE   7 AA;  730 MW;  75B72EA2C73772A0 CRC64;
```

```
  Query Match             29.2%;  Score 14;  DB 13;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 CLN 8
            ||:
Db        5 CLS 7


RESULT   8
Q9X3K1
ID   Q9X3K1      PRELIMINARY;      PRT;     8 AA.
AC   Q9X3K1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CYTOCHROME B (FRAGMENT).
GN   PETB.
OS   Prochlorococcus sp.
OC   Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC   Prochlorococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Urbach E., Chisholm S.W.;
RT   "Genetic diversity in Prochlorococcus populations flow cytometrically
RT   sorted from the Sargasso Sea and Gulf Stream.";
RL   Limnol. Oceanog. 43:1615-1630(1998).
DR   EMBL; AF070193; AAD23233.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  799 MW;  10376865B72866D3 CRC64;


  Query Match             29.2%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSG 4
            |||
Db        4 LSG 6


RESULT   9
P82079
ID   P82079      PRELIMINARY;      PRT;     8 AA.
AC   P82079;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   DYNASTIN 1.
OS   Limnodynastes interioris (Giant banjo frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Myobatrachidae;
OC   Limnodynastes.
RN   [1]
RP   SEQUENCE, AND MASS SPECTROMETRY.
RC   TISSUE=TIBIAL GLAND;
RA   Raftery M.J., Bradford A.M., Bowie J.H., Wallace J.C., Tyler M.J.;
RT   "Peptides from Australian frogs. The structure of the dynastins from
RT   the banjo frogs Limnodynastes interioras, Limnodynastes dumerilii and
RT   Limnodynastes terraereginae.";
RL   Aust. J. Chem. 46:833-842(1993).
CC   -!- MASS SPECTROMETRY: MW=729; METHOD=FAB.
KW   Amphibian skin.
SQ   SEQUENCE   8 AA;  729 MW;  7C28772865B72728 CRC64;


  Query Match             29.2%;  Score 14;  DB 13;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSG 4
            |||
```

```
Db       3 LSG 5


RESULT  10
Q9TRS0
ID   Q9TRS0      PRELIMINARY;      PRT;     9 AA.
AC   Q9TRS0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CALCYCLIN-ASSOCIATED PROTEIN, CAP50=CA2+/PHOSPHOLIPID-BINDING PROTEIN
DE   L-7 FRAGMENT.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92250478.
RA   Tokumitsu H., Mizutani A., Minami H., Kobayashi R., Hidaka H.;
RT   "A calcyclin-associated protein is a newly identified member of the
RT   Ca2+/phospholipid-binding proteins, annexin family.";
RL   J. Biol. Chem. 267:8919-8924(1992).
SQ   SEQUENCE   9 AA;  1010 MW;  64E419C44865B72B CRC64;


  Query Match             29.2%;  Score 14;  DB 6;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           |||
Db       3 LSG 5


RESULT  11
O35953
ID   O35953      PRELIMINARY;      PRT;     9 AA.
AC   O35953;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   SODIUM CHANNEL, VOLTAGE-GATED, TYPE VIII, ALPHA POLYPEPTIDE
DE   (TRUNCATED VOLTAGE-GATED SODIUM CHANNEL ALPHA SUBUNIT) (FRAGMENT).
GN   SCN8A.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIII;
RX   MEDLINE; 97442476.
RA   Plummer N.W., McBurney M.W., Meisler M.H.;
RT   "Alternative splicing of the sodium channel SCN8A predicts a truncated
RT   two-domain protein in fetal brain and non-neuronal cells.";
RL   J. Biol. Chem. 272:24008-24015(1997).
DR   EMBL; U97672; AAB80914.1; -.
DR   MGD; MGI:103169; Scn8a.
KW   Ionic channel.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  898 MW;  22D92865B735B737 CRC64;


  Query Match             29.2%;  Score 14;  DB 11;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           |||
Db       5 LSG 7


RESULT  12
Q08720
ID   Q08720      PRELIMINARY;      PRT;     6 AA.
AC   Q08720;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   Y PROTEIN (FRAGMENT).
GN   CREB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93010691.
RA   Waeber G., Habener J.F.;
RT   "Novel testis germ cell-specific transcript of the CREB gene contains
RT   an alternatively spliced exon with multiple in-frame stop codons.";
RL   Endocrinology 131:2010-2015(1992).
DR   EMBL; X68994; CAA48780.1; -.
FT   NON_TER       1      1
FT   NON_TER       6      6
SQ   SEQUENCE   6 AA;  695 MW;  67272EB9C735D000 CRC64;


  Query Match             27.1%;  Score 13;  DB 4;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 CL 7
           ||
Db       4 CL 5


RESULT  13
Q9Y4X6
ID   Q9Y4X6      PRELIMINARY;      PRT;     8 AA.
AC   Q9Y4X6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   NUCLEAR LIM INTERACTOR (FRAGMENT).
GN   NLI.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Drechsler M., Schumacher V., Friedrich S., Wildhardt G., Giesler S.,
RA   Schroth A., Bodem J., Royer-Pokora B.;
RT   "Genomic structure, alternative transcripts and chromosomal
RT   localization of the human LIM domain binding protein 1 gene
RT   LDB1/NLI.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ243097; CAB45408.1; -.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  767 MW;  EE6EBDDEB862D5B6 CRC64;


  Query Match             27.1%;  Score 13;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 AC 6
           ||
Db       6 AC 7


RESULT  14
O02831
ID   O02831      PRELIMINARY;      PRT;     8 AA.
AC   O02831;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
```

```
DE   PRO ALPHA 1 TYPE III COLLAGEN PROTEIN (FRAGMENT).
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96377339.
RA   Metsaranta M., Kujala U.M., Pelliniemi L., Osterman H., Aho H.,
RA   Vuorio E.;
RT   "Evidence for insufficient chondrocytic differentiation during repair
RT   of full-thickness defects of articular cartilage.";
RL   Matrix Biol. 15:39-47(1996).
DR   EMBL; S83371; AAD14433.1; -.
FT   NON_TER      1      1
SQ   SEQUENCE   8 AA;  1028 MW;  B859C7272EA77371 CRC64;


  Query Match             27.1%;  Score 13;  DB 6;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 CL 7
           ||
Db       4 CL 5


RESULT  15
Q47063
ID   Q47063      PRELIMINARY;      PRT;     9 AA.
AC   Q47063;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   URF 1.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W;
RX   MEDLINE; 85215599.
RA   Takagi J.S., Ida N., Tokushige M., Sakamoto H., Shimura Y.;
RT   "Cloning and nucleotide sequence of the aspartase gene of Escherichia
RT   coli W.";
RL   Nucleic Acids Res. 13:2063-2074(1985).
DR   EMBL; X02307; CAA26175.1; -.
SQ   SEQUENCE   9 AA;  1061 MW;  9DE21EA5B9C72EA1 CRC64;


  Query Match             27.1%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 CL 7
           ||
Db       3 CL 4


RESULT  16
Q99887
ID   Q99887      PRELIMINARY;      PRT;     9 AA.
AC   Q99887;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   11 <BETA>-HSD2 PROTEIN (FRAGMENT).
GN   11 <BETA>-HSD2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96133030.
RA   Stewart P.M., Krozowski Z.S., Gupta A., Milford D.V., Howie A.J.,
RA   Sheppard M.C., Whorwood C.B.;
RT   "Hypertension in the syndrome of apparent mineralocorticoid excess due
RT   to mutation of the 11 beta-hydroxysteroid dehydrogenase type 2 gene.";
RL   Lancet 347:88-91(1996).
DR   EMBL; S80133; AAD14324.1; -.
FT   NON_TER      1      1
FT   NON_TER      9      9
SQ   SEQUENCE   9 AA;  1020 MW;  CEFC2EB1F5B059C9 CRC64;

  Query Match             27.1%;  Score 13;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 CL 7
           ||
Db       7 CL 8


RESULT  17
Q9QZA8
ID   Q9QZA8      PRELIMINARY;      PRT;     9 AA.
AC   Q9QZA8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   C-TYPE LECTIN DCL1 (FRAGMENT).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Gorski K., Huang X., Tseng S.-Y., Rattis F., Pardoll D., Tsuchiya H.;
RT   "Dendritic cell regulation of DCL1 mRNA expression.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF192526; AAF04843.1; -.
KW   Lectin.
FT   NON_TER      9      9
SQ   SEQUENCE   9 AA;  994 MW;  342161AB172EBAB7 CRC64;


  Query Match             27.1%;  Score 13;  DB 11;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 CL 7
           ||
Db       4 CL 5


RESULT  18
Q69473
ID   Q69473      PRELIMINARY;      PRT;     9 AA.
AC   Q69473;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1996 (TrEMBLrel. 01, Last annotation update)
DE   IMMEDIATE-EARLY TRANSACTIVATOR 110 (FRAGMENT).
GN   ICP0.
OS   human herpesvirus 1.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MP;
RA   Gu W., Huang Q., Hayward G.S.;
RL   J. Biomed. Sci. 2:105-130(1995).
DR   EMBL; U18080; AAA75442.1; -.
FT   NON_TER      1      1
FT   NON_TER      9      9
SQ   SEQUENCE   9 AA;  1029 MW;  797BB867740DDB04 CRC64;
```

```
  Query Match             27.1%;  Score 13;  DB 12;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 CL 7
           ||
Db       8 CL 9

RESULT  19
P77556
ID   P77556      PRELIMINARY;      PRT;     8 AA.
AC   P77556;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   TRAY (FRAGMENT).
GN   TRAY.
OS   Escherichia coli.
OG   Plasmid IncFII R1.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ECOR11;
RX   MEDLINE; 96400908.
RA   Boyd E.F., Hill C.W., Rich S.M., Hartl D.L.;
RT   "Mosaic structure of plasmids from natural populations of Escherichia
RT   coli.";
RL   Genetics 143:1091-1100(1996).
DR   EMBL; U50661; AAC44245.1; -.
DR   EMBL; U50650; AAC44234.1; -.
DR   EMBL; U50651; AAC44235.1; -.
DR   EMBL; U50652; AAC44236.1; -.
DR   EMBL; U50653; AAC44237.1; -.
DR   EMBL; U50654; AAC44238.1; -.
DR   EMBL; U50655; AAC44239.1; -.
DR   EMBL; U50656; AAC44240.1; -.
DR   EMBL; U50657; AAC44241.1; -.
DR   EMBL; U50658; AAC44242.1; -.
DR   EMBL; U50659; AAC44243.1; -.
DR   EMBL; U50660; AAC44244.1; -.
KW   Plasmid.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  834 MW;  D335A5B0544735A1 CRC64;

  Query Match             25.0%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LNL 9
           ||:
Db       3 LNI 5

RESULT  20
Q15901
ID   Q15901      PRELIMINARY;      PRT;     8 AA.
AC   Q15901;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   (CLONE XP7B11B) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   Lee C.C., Yazdani A., Wehnert M., Bailey J., Couch L., Xiong M.,
RA   Coolbaugh M.I., Chinault C.A., Baldini A., Lindsay E.A., Zhao Z.Y.,
```

```
RA   Caskey C.T.H.;
RL   Hum. Mol. Genet. 0:0-0(0).
DR   EMBL; L32080; AAA73891.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  860 MW;  37D72878676729CB CRC64;

  Query Match             25.0%;  Score 12;  DB 4;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           :| |
Db       2 FLPG 5

RESULT  21
Q9UMH9
ID   Q9UMH9      PRELIMINARY;      PRT;     8 AA.
AC   Q9UMH9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RHCE PROTEIN (FRAGMENT).
GN   RHCE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RA   Matassi G., Cherif-Zahar B., Mouro I., Cartron J.P.;
RT   "Characterization of the recombination hot spot involved in the
RT   genomic rearrangement leading to the hybrid D-CE-D gene in the DVI
RT   phenotype.";
RL   Am. J. Hum. Genet. 60:808-817(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE; 90349591.
RA   Cherif-Zahar B., Bloy C., Le Van Kim C., Blanchard D., Bailly P.,
RA   Hermand P., Salmon C., Cartron J.-P., Colin Y.;
RT   "Molecular cloning and protein structure of a human blood group Rh
RT   polypeptide.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:6243-6247(1990).
DR   EMBL; Z97030; CAB09726.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  1049 MW;  C007244691FB5AB1 CRC64;

  Query Match             25.0%;  Score 12;  DB 4;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LNL 9
           :||
Db       5 MNL 7

RESULT  22
Q95213
ID   Q95213      PRELIMINARY;      PRT;     8 AA.
AC   Q95213;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   GERMLINE DH (DF) GENE (FRAGMENT).
GN   DF.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F-I/RGM;
RA   Mage R.G., Chen H.T., Alexander C.B., Chen F.F.;
RL   Mol. Immunol. 0:0-0(0).
DR   EMBL; U62585; AAB18735.1; -.
KW   MHC.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  845 MW;  5CA861B5AB58677B CRC64;

  Query Match             25.0%;  Score 12;  DB 7;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           | :|
Db       4 YSTG 7

RESULT  23
Q90498
ID   Q90498      PRELIMINARY;      PRT;     8 AA.
AC   Q90498;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   MYOGLOBIN (FRAGMENT).
OS   Erythrura gouldiae (Gouldian finch).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Passeridae; Erythrura.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EGG1;
RX   MEDLINE; 98208049.
RA   Heslewood M.M., Elphinstone M.S., Tidemann S.C., Baverstock P.R.;
RT   "Myoglobin intron variation in the Gouldian Finch Erythrura gouldiae
RT   assessed by temperature gradient gel electrophoresis.";
RL   Electrophoresis 19:142-151(1998).
DR   EMBL; U40496; AAC60363.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  890 MW;  C6C1F2C865B046DE CRC64;

  Query Match             25.0%;  Score 12;  DB 13;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       3 ISG 5

RESULT  24
Q91098
ID   Q91098      PRELIMINARY;      PRT;     8 AA.
AC   Q91098;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   MYOGLOBIN (FRAGMENT).
OS   Manorina melanocephala (noisy miner).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Meliphagidae; Manorina.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D02;
RX   MEDLINE; 98208049.
RA   Heslewood M.M., Elphinstone M.S., Tidemann S.C., Baverstock P.R.;
RT   "Myoglobin intron variation in the Gouldian Finch Erythrura gouldiae
RT   assessed by temperature gradient gel electrophoresis.";
RL   Electrophoresis 19:142-151(1998).
DR   EMBL; U40497; AAC60364.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  890 MW;  C6C1F2C865B046DE CRC64;

  Query Match             25.0%;  Score 12;  DB 13;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       3 ISG 5

RESULT  25
P82082
ID   P82082      PRELIMINARY;      PRT;     8 AA.
AC   P82082;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   DYNASTIN 4.
OS   Limnodynastes salmini (Salmin's-striped frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Myobatrachidae;
OC   Limnodynastes.
RN   [1]
RP   SEQUENCE, AND MASS SPECTROMETRY.
RC   TISSUE=SKIN SECRETION;
RA   Bradford A.M., Raftery M.J., Bowie J.H., Wallace J.C., Tyler M.J.;
RT   "Peptides from Australian frogs. The structure of the dynastins from
RT   Limnodynastes salmini and Fletcherin from Limnodynastes fletcheri.";
RL   Aust. J. Chem. 46:1235-1244(1993).
CC   -!- MASS SPECTROMETRY: MW=772; METHOD=FAB.
KW   Amphibian skin.
SQ   SEQUENCE   8 AA;  772 MW;  7B58772455A2C728 CRC64;

  Query Match             25.0%;  Score 12;  DB 13;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 GACLNL 9
           |   ||
Db       1 GLVSNL 6

RESULT  26
P82083
ID   P82083      PRELIMINARY;      PRT;     8 AA.
AC   P82083;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   DYNASTIN 5.
OS   Limnodynastes salmini (Salmin's-striped frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Myobatrachidae;
OC   Limnodynastes.
RN   [1]
RP   SEQUENCE, AND MASS SPECTROMETRY.
RC   TISSUE=SKIN SECRETION;
RA   Bradford A.M., Raftery M.J., Bowie J.H., Wallace J.C., Tyler M.J.;
RT   "Peptides from Australian frogs. The structure of the dynastins from
RT   Limnodynastes salmini and Fletcherin from Limnodynastes fletcheri.";
RL   Aust. J. Chem. 46:1235-1244(1993).
CC   -!- MASS SPECTROMETRY: MW=786; METHOD=FAB.
KW   Amphibian skin.
SQ   SEQUENCE   8 AA;  786 MW;  7B58772455B05728 CRC64;
```

```
  Query Match             25.0%;  Score 12;  DB 13;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 GACLNL 9
           |   ||
Db       1 GLISNL 6

RESULT  27
Q9R635
ID   Q9R635      PRELIMINARY;      PRT;     9 AA.
AC   Q9R635;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR OUTER MEMBRANE PROTEIN VARIABLE DOMAIN IV, MOMP VD IV.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92040090.
RA   Peterson E.M., Cheng X., Markoff B.A., Fielder T.J., de la Maza L.M.;
RT   "Functional and structural mapping of Chlamydia trachomatis species-
RT   specific major outer membrane protein epitopes by use of neutralizing
RT   monoclonal antibodies.";
RL   Infect. Immun. 59:4147-4153(1991).
SQ   SEQUENCE   9 AA;  976 MW;  9C61B041B7645361 CRC64;

  Query Match             25.0%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       7 ISG 9

RESULT  28
Q9UCN5
ID   Q9UCN5      PRELIMINARY;      PRT;     9 AA.
AC   Q9UCN5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CELL-SURFACE HEPARIN/HEPARANSULFATE-BINDING PROTEIN PEPTIDE 2
DE   (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92291065.
RA   Raboudi N., Julian J., Rohde L.H., Carson D.D.;
RT   "Identification of cell-surface heparin/heparan sulfate-binding
RT   proteins of a human uterine epithelial cell line (RL95).";
RL   J. Biol. Chem. 267:11930-11939(1992).
SQ   SEQUENCE   9 AA;  1008 MW;  CB56D0544732C732 CRC64;

  Query Match             25.0%;  Score 12;  DB 4;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LNL 9
           ||:
Db       5 LNI 7

RESULT  29
Q27396
ID   Q27396      PRELIMINARY;      PRT;     9 AA.
```

```
AC   Q27396;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   RHOPTRY ASSOCIATED PROTEIN 1.
GN   RAP-1.
OS   Babesia bovis.
OC   Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MO7;
RA   Suarez C.E., Palmer G.H., Hines S.A., McElwain T.F.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L77326; AAA96415.1; -.
SQ   SEQUENCE   9 AA;  931 MW;  6D5BD2C865B05044 CRC64;

  Query Match             25.0%;  Score 12;  DB 5;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       4 ISG 6

RESULT  30
Q28112
ID   Q28112      PRELIMINARY;      PRT;     9 AA.
AC   Q28112;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   6-PHOSPHOFRUCTO-2-KINASE/FRUCTOSE-2, 6-BISPHOSPHATASE (FRAGMENT).
GN   GENE B.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93387464.
RA   Vidal H., Crepin K.M., Rider M.H., Hue L., Rousseau G.G.;
RT   "Cloning and expression of novel isoforms of 6-phosphofructo-2-
RT   kinase/fructose-2,6-bisphosphatase from bovine heart.";
RL   FEBS Lett. 330:329-333(1993).
DR   EMBL; X74564; CAA52652.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  837 MW;  859CA5BDC7644865 CRC64;

  Query Match             25.0%;  Score 12;  DB 6;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       1 MSG 3

RESULT  31
Q9TRW2
ID   Q9TRW2      PRELIMINARY;      PRT;     9 AA.
AC   Q9TRW2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CALDESMON=PHOSPHORYLATION SITE.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE.
```

```
RX   MEDLINE; 91378498.
RA   Ikebe M., Hornick T.;
RT   "Determination of the phosphorylation sites of smooth muscle caldesmon
RT   by protein kinase C.";
RL   Arch. Biochem. Biophys. 288:538-542(1991).
SQ   SEQUENCE   9 AA;  1018 MW;  8C901B10533735A5 CRC64;


  Query Match             25.0%;  Score 12;  DB 6;  Length 9;
  Best Local Similarity   33.3%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        4 GACLNL 9
            |: | :
Db        1 GSSLKI 6


RESULT  32
Q31653
ID   Q31653      PRELIMINARY;      PRT;     9 AA.
AC   Q31653;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CYTOCHROME B (FRAGMENT).
GN   CYTOCHROME B.
OS   Anser caerulescens (Goose).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anser.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94141942.
RA   Quinn T.W., Wilson A.C.;
RT   "Sequence evolution in and around the mitochondrial control region in
RT   birds.";
RL   J. Mol. Evol. 37:417-425(1993).
DR   EMBL; X77190; CAA54411.1; -.
KW   Mitochondrion.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1095 MW;  4751472693344B17 CRC64;


  Query Match             25.0%;  Score 12;  DB 8;  Length 9;
  Best Local Similarity   42.9%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        2 LSGACLN 8
            |    ||
Db        2 LENKMLN 8


RESULT  33
Q67113
ID   Q67113      PRELIMINARY;      PRT;     7 AA.
AC   Q67113;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   (UDORN/72) HEMAGGLUTININ (SEG 4) CDNA, 3' END (FRAGMENT).
OS   Influenza A virus.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza virus A and B group.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 81001892.
RA   Dhar R., Chanock R.M., Lai C.J.;
RT   "Nonviral oligonucleotides at the 5' terminus of cytoplasmic influenza
RT   viral mRNA deduced from cloned complete genomic sequences.";
RL   Cell 21:495-500(1980).
DR   EMBL; M25045; AAA43202.1; -.
KW   Hemagglutinin.
FT   NON_TER       1      1
```

```
SQ   SEQUENCE   7 AA;  834 MW;  605EB0544EA40030 CRC64;

  Query Match             22.9%;  Score 11;  DB 12;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    1;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 CL 7
            |:
Db        6 CI 7


RESULT  34
Q9YQ10
ID   Q9YQ10      PRELIMINARY;      PRT;     7 AA.
AC   Q9YQ10;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HYPOTHETICAL FUSION PROTEIN.
OS   porcine transmissible gastroenteritis virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99099045.
RA   Izeta A., Smerdou C., Alonso S., Penzes Z., Mendez A., Plana-Duran J.,
RA   Enjuanes L.;
RT   "Replication and packaging of transmissible gastroenteritis
RT   coronavirus-derived synthetic minigenomes.";
RL   J. Virol. 73:1535-1545(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95159435.
RA   Eleouet J., Rasschaert D., Lambert P., Levy L., Vende P., Laude H.;
RT   "Complete sequence (20 kilobases) of the polyprotein-encoding gene 1
RT   of transmissible gastroenteritis virus.";
RL   Virology 206:817-822(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88078100.
RA   Rasschaert D., Gelfi J., Laude H.;
RT   "Enteric coronavirus TGEV: partial sequence of the genomic RNA, its
RT   organization and expression.";
RL   Biochimie 69:591-600(1987).
DR   EMBL; AJ011482; CAA09625.1; -.
SQ   SEQUENCE   7 AA;  927 MW;  69D6D7273B5726F0 CRC64;


  Query Match             22.9%;  Score 11;  DB 12;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YL 2
            ||
Db        3 YL 4


RESULT  35
O32560
ID   O32560      PRELIMINARY;      PRT;     8 AA.
AC   O32560;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   PROPIONATE KINASE (FRAGMENT).
GN   TDCD.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W3110;
```

```
RA   Hesslinger C., Sawers G.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=W3110;
RA   Hesslinger C., Fairhurst S.A., Sawers G.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ001620; CAA04875.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  1000 MW;  3A505EB044140DC4 CRC64;

  Query Match             22.9%;  Score 11;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    1;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 CL 7
          |:
Db      6 CI 7


Search completed: December 16, 2000, 04:22:13
Job time: 4607 sec
```

```
                        GenCore version 4.5
              Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 04:23:31 ; Search time 68.03 Seconds
                                               (without alignments)
                                               4.227 Million cell updates/sec

Title:           US-09-529-121-5
Perfect score:   48
Sequence:        1 YLSGACLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters:      213

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       SwissProt_39:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                     %
Result             Query
  No.    Score  Match Length DB  ID                         Description
-------------------------------------------------------------------------------
    1       17   35.4      8  1  CPD1_ENTFA                 P13269 enterococcu
    2       15   31.2      9  1  SAP_STOVA                  P24047 stomopneute
    3       13   27.1      9  1  DNF1_LOCMI                 P16339 locusta mig
    4       13   27.1      9  1  ISOT_CYPCA                 P42993 cyprinus ca
    5       13   27.1      9  1  OXYT_RAJCL                 P42994 raja clavat
    6       12   25.0      9  1  PGLR_DIAAB                 P81179 diaprepes a
    7       11   22.9      4  1  FAR3_HIRME                 P42562 hirudo medi
    8       11   22.9      5  1  PRCT_PERAM                 P01373 periplaneta
    9       11   22.9      7  1  FAR2_ASCSU                 P31890 ascaris suu
   10       11   22.9      7  1  GFRP_MOUSE                 P99025 mus musculu
   11       11   22.9      8  1  CAD1_ENTFA                 P13268 enterococcu
   12       11   22.9      9  1  CONO_CONST                 P05487 conus stria
   13       11   22.9      9  1  MOSF_CLYJA                 P19853 clypeaster
   14       11   22.9      9  1  OXYA_SQUAC                 P42999 squalus aca
   15       10   20.8      5  1  UXA4_CHLTR                 P38005 chlamydia t
   16       10   20.8      6  1  CIP2_MYTED                 P13737 mytilus edu
   17       10   20.8      6  1  TMOF_SARBU                 P41495 sarcophaga
   18       10   20.8      7  1  ALL2_CARMA                 P81805 carcinus ma
   19       10   20.8      7  1  ALL3_CARMA                 P81806 carcinus ma
   20       10   20.8      7  1  ALL4_CARMA                 P81807 carcinus ma
   21       10   20.8      7  1  ALL5_CARMA                 P81808 carcinus ma
   22       10   20.8      7  1  UN06_PINPS                 P81675 pinus pinas
   23       10   20.8      8  1  AKH_MELML                  P25423 melolontha
   24       10   20.8      8  1  AL12_CARMA                 P81815 carcinus ma
   25       10   20.8      8  1  AL17_CARMA                 P81820 carcinus ma
   26       10   20.8      8  1  ALL7_CARMA                 P81809 carcinus ma
   27       10   20.8      8  1  ALL8_CARMA                 P81811 carcinus ma
   28       10   20.8      8  1  ALL9_CARMA                 P81812 carcinus ma
   29       10   20.8      8  1  FAR8_CALVO                 P41863 calliphora
   30       10   20.8      8  1  GLUR_HUMAN                 P02729 homo sapien
   31       10   20.8      8  1  HTF_TENMO                  P25419 tenebrio mo
   32       10   20.8      8  1  LCK5_LEUMA                 P19987 leucophaea
   33       10   20.8      8  1  LCK8_LEUMA                 P19990 leucophaea
   34       10   20.8      8  1  RPCH_PANBO                 P08939 pandalus bo
   35       10   20.8      8  1  UF06_MOUSE                 P38644 mus musculu
   36       10   20.8      8  1  VGLG_HSV2B                 P81780 herpes simp
   37       10   20.8      9  1  AL10_CARMA                 P81813 carcinus ma
   38       10   20.8      9  1  AL11_CARMA                 P81814 carcinus ma
   39       10   20.8      9  1  DSIP_RABIT                 P01158 oryctolagus
   40       10   20.8      9  1  FAR5_ASCSU                 P43170 ascaris suu
   41       10   20.8      9  1  FAR6_CALVO                 P41861 calliphora
   42       10   20.8      9  1  FIBB_MACFU                 P19345 macaca fusc
   43       10   20.8      9  1  LMIP_LOCMI                 P31799 locusta mig
   44       10   20.8      9  1  MOSH_CLYJA                 P19852 clypeaster
   45       10   20.8      9  1  OXYA_SCYCA                 P42996 scyliorhinu
   46       10   20.8      9  1  OXYT_BUFRE                 P42995 bufo regula
   47       10   20.8      9  1  OXYT_RABIT                 P32878 oryctolagus
   48       10   20.8      9  1  TAL3_PICJA                 P17441 pichia jadi
   49       10   20.8      9  1  TKL1_LOCMI                 P16223 locusta mig
   50       10   20.8      9  1  TRP4_LEUMA                 P81736 leucophaea
   51       10   20.8      9  1  ULAH_HUMAN                 P31934 homo sapien
   52       10   20.8      9  1  UPA6_HUMAN                 P30092 homo sapien
   53        9   18.8      4  1  FAR4_HIRME                 P42563 hirudo medi
   54        9   18.8      5  1  AL14_CARMA                 P81817 carcinus ma
   55        9   18.8      5  1  FARP_ARTTR                 P41853 artioposthi
   56        9   18.8      7  1  FAR3_HAECO                 P81298 haemonchus
   57        9   18.8      7  1  FAR3_PANRE                 P41874 panagrellus
   58        9   18.8      7  1  FAR5_HIRME                 P42564 hirudo medi
   59        9   18.8      8  1  ACT_CARMA                  P80709 carcinus ma
   60        9   18.8      8  1  AL15_CARMA                 P81818 carcinus ma
   61        9   18.8      8  1  AL16_CARMA                 P81819 carcinus ma
   62        9   18.8      8  1  AL18_CARMA                 P81821 carcinus ma
   63        9   18.8      8  1  ALL3_CYDPO                 P82154 cydia pomon
   64        9   18.8      8  1  ALL4_CALVO                 P41840 calliphora
   65        9   18.8      8  1  ALL4_CYDPO                 P82155 cydia pomon
   66        9   18.8      8  1  ANG2_BOTJA                 Q10582 bothrops ja
   67        9   18.8      8  1  CCKN_MACEU                 P30369 macropus eu
   68        9   18.8      8  1  LPMS_STAEP                 P23211 staphylococ
   69        9   18.8      9  1  CCAP_CARMA                 P38556 carcinus ma
   70        9   18.8      9  1  CONO_CONGE                 P05486 conus geogr
   71        9   18.8      9  1  FABF_RHIME                 P56902 rhizobium m
   72        9   18.8      9  1  FLA2_TREHY                 P80159 treponema h
   73        9   18.8      9  1  LPCA_STAAU                 P36884 staphylococ
   74        9   18.8      9  1  MGMT_BOVIN                 P29177 bos taurus
   75        9   18.8      9  1  OXYF_SCYCA                 P42997 scyliorhinu


                                  ALIGNMENTS


RESULT   1
CPD1_ENTFA
ID   CPD1_ENTFA     STANDARD;      PRT;     8 AA.
AC   P13269;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   SEX PHEROMONE CPD1.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 85040388.
RA   Suzuki A., Mori M., Sagakami Y., Isogai A., Fujino M., Kitada C.,
RA   Craig R.A., Clewell D.B.;
RT   "Isolation and structure of bacterial sex pheromone, cPD1.";
RL   Science 226:849-850(1984).
CC   -!- FUNCTION: CPD1 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE
CC       BACTERIOCIN PLASMID PPD1.
KW   Pheromone.
SQ   SEQUENCE   8 AA;  913 MW;  8665B729C682C729 CRC64;

  Query Match             35.4%;  Score 17;  DB 1;  Length 8;
  Best Local Similarity   75.0%;  Pred. No. 8.8e+04;
```

```
  Matches     3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           :|||
Db       5 FLSG 8


RESULT   2
SAP_STOVA
ID   SAP_STOVA      STANDARD;      PRT;     9 AA.
AC   P24047;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   01-MAR-1992 (Rel. 21, Last annotation update)
DE   SPERM-ACTIVATING PEPTIDE (SAP).
OS   Stomopneutes variolaris (Sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Diadematacea; Phymosomatoida; Stomechinidae;
OC   Stomopneustes.
RN   [1]
RP   SEQUENCE, AND DISULFIDE BOND.
RC   TISSUE=EGG JELLY;
RX   MEDLINE; 92097763.
RA   Yoshimo K.-I., Takao T., Shimonishi Y., Suzuki N.;
RT   "Determination of the amino acid sequence of an intramolecular
RT   disulfide linkage-containing sperm-activating peptide by tandem mass
RT   spectrometry.";
RL   FEBS Lett. 294:179-182(1991).
CC   -!- FUNCTION: CAUSE STIMULATION OF SPERM RESPIRATION AND MOTILITY
CC       THROUGH INTRACELLULAR ALKALINIZATION, TRANSIENT ELEVATIONS OF
CC       CAMP, CGMP AND CLACIUM LEVELS IN SPERM CELLS, AND TRANSIENT
CC       ACTIVATION AND SUBSEQUENT INACTIVATION OF THE MEMBRANE FORM OF
CC       GUANYLATE CYCLASE.
DR   PIR; S19329; S19329.
FT   DISULFID      3      8
SQ   SEQUENCE   9 AA;  1010 MW;  C469B3387B076EB9 CRC64;


  Query Match             31.2%;  Score 15;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches     2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GACL 7
           | |:
Db       6 GKCV 9


RESULT   3
DNF1_LOCMI
ID   DNF1_LOCMI     STANDARD;      PRT;     9 AA.
AC   P16339;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   LOCUPRESSIN (DIURETIC NEUROPEPTIDE F1/F2).
OS   Locusta migratoria (Migratory locust).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC   Acridomorpha; Acridoidea; Acrididae; Locusta.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=SUBOESOPHAGEAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 88077077.
RA   Proux J.P., Miller C.A., Li J.P., Carney R.L., Girardie A.,
RA   Delaage M., Schooley D.A.;
RT   "Identification of an arginine vasopressin-like diuretic hormone from
RT   Locusta migratoria.";
RL   Biochem. Biophys. Res. Commun. 149:180-186(1987).
CC   -!- FUNCTION: DIURETIC HORMONE.
CC   -!- SUBUNIT: F2 IS AN ANTIPARALLEL DISULFIDE LINKED DIMER OF F1.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   PIR; A29477; A29477.
DR   INTERPRO; IPR000981; -.
```

```
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Neuropeptide; Amidation.
FT   DISULFID      1      6       IN F1.
FT   DISULFID      1      1       INTERCHAIN (WITH C-6') (IN F2).
FT   DISULFID      6      6       INTERCHAIN (WITH C-1') (IN F2).
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  976 MW;  56EB176EB451A057 CRC64;

  Query Match             27.1%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches     2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 CL 7
           ||
Db       1 CL 2


RESULT   4
ISOT_CYPCA
ID   ISOT_CYPCA     STANDARD;      PRT;     9 AA.
AC   P42993;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ISOTOCIN.
OS   Cyprinus carpio (Common carp).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PITUITARY;
RA   Acher R., Chauvet J., Chauvet M.-T., Crepy D.;
RT   "Characterization of neurohypophyseal hormones from a fresh water bony
RT   fish, the carp (Cyprinus carpio). Comparison with hormones from sea
RT   water bony fishs.";
RL   Comp. Biochem. Physiol. 14:245-254(1965).
CC   -!- FUNCTION: ANTIDIURETIC HORMONE.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   PIR; A61364; A61364.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  969 MW;  17FF476EB455B04B CRC64;


  Query Match             27.1%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches     2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLS 3
           |:|
Db       2 YIS 4


RESULT   5
OXYT_RAJCL
ID   OXYT_RAJCL     STANDARD;      PRT;     9 AA.
AC   P42994;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   GLUMITOCIN.
OS   Raja clavata (Thornback ray).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Hypnosqualea; Pristiorajea; Batoidea;
OC   Rajiformes; Rajidae; Raja.
RN   [1]
```

```
RP   SEQUENCE.
RX   MEDLINE; 66123415.
RA   Acher R., Chauvet J., Chauvet M.-T., Crepy D.;
RT   "Phylogeny of neurophyophyseal peptides: isolation of a new hormone,
RT   glumitocin (Ser 4-Gln 8-ocytocin) present in a cartilaginous fish,
RT   the ray (Raia clavata).";
RL   Biochim. Biophys. Acta 107:393-396(1965).
CC   -!- FUNCTION: ANTIDIURETIC HORMONE.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  984 MW;  17E9C76EB455B04B CRC64;


  Query Match             27.1%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLS 3
            |:|
Db        2 YIS 4


RESULT   6
PGLR_DIAAB
ID   PGLR_DIAAB     STANDARD;      PRT;     9 AA.
AC   P81179;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ENDO-POLYGALACTURONASE (PG) (EC 3.2.1.15) (FRAGMENT).
OS   Diaprepes abbreviatus (Sugarcane rootstalk borer weevil).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC   Cucujiformia; Curculionidae; Entiminae; Entimini; Diaprepes.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LARVAL GUT;
RA   Doostdar H., McCollum T.G., Mayer R.T.;
RT   "Purification and characterization of an endo-polygalacturonase from
RT   the gut of West Indies sugarcane rootstalk borer weevil (Diaprepes
RT   abbreviatus L.) larvae.";
RL   Comp. Biochem. Physiol. 118B:861-867(1997).
CC   -!- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-ALPHA-D-
CC       GALACTOSIDURONIC LINKAGES IN PECTATE AND OTHER GALACTURONANS.
CC   -!- INDUCTION: INHIBITED BY CITRUS PGIP.
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS PROTEIN IS:
CC       9.4, ITS MW IS: 44.5 KDA.
CC   -!- SIMILARITY: WEAK TO OTHER POLYGALACTURONASES.
KW   Hydrolase; Glycosidase; Cell wall.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1041 MW;  1F49087042DB41BB CRC64;


  Query Match             25.0%;  Score 12;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSG 4
            |: |
Db        4 YVIG 7


RESULT   7
FAR3_HIRME
ID   FAR3_HIRME     STANDARD;      PRT;     4 AA.
AC   P42562;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE YLRF-AMIDE.
OS   Hirudo medicinalis (Medicinal leech).
OC   Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;
OC   Arynchobdellida; Hirudiniformes; Hirudinidae; Hirudo.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92195954.
RA   Evans B.D., Pohl J., Kartsonis M.A., Calabrese R.L.;
RT   "Identification of RFamide neuropeptides in the medicinal leech.";
RL   Peptides 12:897-908(1991).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       4      4       AMIDATION.
SQ   SEQUENCE   4 AA;  598 MW;  69D4073B30000000 CRC64;


  Query Match             22.9%;  Score 11;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YL 2
            ||
Db        1 YL 2


RESULT   8
PRCT_PERAM
ID   PRCT_PERAM     STANDARD;      PRT;     5 AA.
AC   P01373;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   PROCTOLIN.
OS   Periplaneta americana (American cockroach),
OS   Limulus polyphemus (Atlantic horseshoe crab), and
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blattoidea; Blattidae; Periplaneta.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=P.AMERICANA;
RX   MEDLINE; 76074708.
RA   Starratt A.N., Brown B.E.;
RT   "Structure of the pentapeptide proctolin, a proposed neurotransmitter
RT   in insects.";
RL   Life Sci. 17:1253-1256(1975).
RN   [2]
RP   BIOLOGICAL SOURCE.
RC   SPECIES=P.AMERICANA;
RX   MEDLINE; 81225865.
RA   O'Shea M., Adams M.E.;
RT   "Pentapeptide (proctolin) associated with an identified neuron.";
RL   Science 213:567-569(1981).
RN   [3]
RP   SEQUENCE.
RC   SPECIES=L.POLYPHEMUS;
RX   MEDLINE; 90287800.
RA   Groome J.R., Tillinghast E.K., Townley M.A., Vetrovs A.,
RA   Watson W.H. III, Hunt D.F., Griffin P.R., Alexander J.E.,
RA   Shabanowitz J.;
RT   "Identification of proctolin in the central nervous system of the
RT   horseshoe crab, Limulus polyphemus.";
RL   Peptides 11:205-211(1990).
RN   [4]
RP   SEQUENCE.
RC   SPECIES=C.MAENAS;
RX   MEDLINE; 86232789.
RA   Stangier J., Dircksen H., Keller R.;
RT   "Identification and immunocytochemical localization of proctolin in
RT   pericardial organs of the shore crab, Carcinus maenas.";
```

```
RL   Peptides 7:67-72(1986).
CC   -!- FUNCTION: STIMULATES CARDIAC OUTPUT AND HINDGUT MOTILITY,
CC       MODULATES VISCERAL AND SKELETAL MUSCLE IN MANY ARTHROPODS.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE LATERAL WHITE NEURONS AND IN
CC       THE CRAB PERICARDIAL ORGANS.
DR   PIR; A01644; HOROHA.
DR   PIR; A60411; A60411.
KW   Neuropeptide.
SQ   SEQUENCE   5 AA;  649 MW;  71B7673B44600000 CRC64;

  Query Match             22.9%;  Score 11;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YL 2
           ||
Db       2 YL 3


RESULT   9
FAR2_ASCSU
ID   FAR2_ASCSU     STANDARD;      PRT;     7 AA.
AC   P31890;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE AF2.
OS   Ascaris suum (Pig roundworm) (Ascaris lumbricoides), and
OS   Panagrellus redivivus.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC   Ascarididae; Ascaris.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=A.SUUM;
RX   MEDLINE; 93324431.
RA   Cowden C., Stretton A.O.W.;
RT   "AF2, an Ascaris neuropeptide: isolation, sequence, and bioactivity.";
RL   Peptides 14:423-430(1993).
RN   [2]
RP   SEQUENCE.
RC   SPECIES=P.REDIVIVUS;
RX   MEDLINE; 95060998.
RA   Maule A.G., Shaw C., Bowman J.W.;
RT   "The FMRFamide-like neuropeptide AF2 (Ascaris suum) is present in the
RT   free-living nematode, Panagrellus redivivus (Nematoda, Rhabditida).";
RL   Parasitology 109:351-356(1994).
CC   -!- FUNCTION: HAS EFFECTS ON MUSCLE TENSION.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE NERVE CORDS AND A VARIETY OF
CC       GANGLIA PARTICULARLY IN THE ANTERIOR REGIONS.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  992 MW;  69D4073B5B11E350 CRC64;

  Query Match             22.9%;  Score 11;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YL 2
           ||
Db       4 YL 5


RESULT  10
GFRP_MOUSE
ID   GFRP_MOUSE     STANDARD;      PRT;     7 AA.
AC   P99025;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GTP CYCLOHYDROLASE I FEEDBACK REGULATORY PROTEIN (P35) (FRAGMENT).
GN   GFRP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LIVER;
RA   Sanchez J.-C., Rouge V., Frutiger S., Hughes G., Yan J.X.,
RA   Hoogland C., Appel R.D., Binz P.-A., Hochstrasser D.F.,
RA   Cowthorne M.;
RL   Submitted (AUG-1998) to the SWISS-PROT data bank.
CC   -!- FUNCTION: MEDIATES TETRAHYDROBIOPTERIN INHIBITION OF GTP
CC       CYCLOHYDROLASE I. THIS INHIBITION IS REVERSED BY L-PHENYLALANINE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
DR   SWISS-2DPAGE; P99025; MOUSE.
FT   INIT_MET      0      0
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  806 MW;  71B5B057273B4700 CRC64;

  Query Match             22.9%;  Score 11;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YL 2
           ||
Db       2 YL 3


RESULT  11
CAD1_ENTFA
ID   CAD1_ENTFA     STANDARD;      PRT;     8 AA.
AC   P13268;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   SEX PHEROMONE CAD1.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 85051889.
RA   Mori M., Sagakami Y., Narita M., Isogai A., Fujino M., Kitada C.,
RA   Craig R.A., Clewell D.B., Suzuki A.;
RT   "Isolation and structure of the bacterial sex pheromone, cAD1, that
RT   induces plasmid transfer in Streptococcus faecalis.";
RL   FEBS Lett. 178:97-100(1984).
CC   -!- FUNCTION: CAD1 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE
CC       HEMOLYSIN PLASMID PAD1.
KW   Pheromone.
SQ   SEQUENCE   8 AA;  819 MW;  047DD732C735B9C7 CRC64;

  Query Match             22.9%;  Score 11;  DB 1;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           |:|
Db       6 LAG 8


RESULT  12
CONO_CONST
ID   CONO_CONST     STANDARD;      PRT;     9 AA.
AC   P05487;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ARG-CONOPRESSIN S.
```

```
OS   Conus striatus (Striated cone).
OC   Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
OC   Neogastropoda; Conoidea; Conidae; Conus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 88058932.
RA   Cruz L.J., de Santos V., Zafaralla G.C., Ramilo C.A., Zeikus R.D.,
RA   Gray W.R., Olivera B.M.;
RT   "Invertebrate vasopressin/oxytocin homologs. Characterization of
RT   peptides from Conus geographus and Conus straitus venoms.";
RL   J. Biol. Chem. 262:15821-15824(1987).
RN   [2]
RP   REVIEW.
RX   MEDLINE; 89024586.
RA   Gray W.R., Olivera B.M., Cruz L.J.;
RT   "Peptide toxins from venomous Conus snails.";
RL   Annu. Rev. Biochem. 57:665-700(1988).
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   PIR; B28495; B28495.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1031 MW;  17EB176EB4540050 CRC64;


  Query Match             22.9%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    1;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 CL 7
            |:
Db        1 CI 2


RESULT 13
MOSF_CLYJA
ID   MOSF_CLYJA     STANDARD;      PRT;      9 AA.
AC   P19853;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   [PHE-6]-MOSACT.
OS   Clypeaster japonicus (Sand dollar).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Gnathostomata; Clypeasteroida;
OC   Clypeasteridae; Clypeaster.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=EGG JELLY;
RA   Suzuki N., Kurita M., Yoshino K.I., Kajiura H., Nomura K.,
RA   Yamaguchi M.;
RT   "Purification and structure of mosact and its derivatives from the
RT   egg jelly of the sea urchin Clypeaster japonicus.";
RL   Zool. Sci. 4:649-656(1987).
CC   -!- FUNCTION: STIMULATES SPERM RESPIRATION AND MOTILITY.
DR   PIR; JN0027; JN0027.
SQ   SEQUENCE   9 AA;  924 MW;  93245729CDC5BAB5 CRC64;


  Query Match             22.9%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSG 4
            :| |
Db        6 FLIG 9


RESULT 14
OXYA_SQUAC
ID   OXYA_SQUAC     STANDARD;      PRT;      9 AA.
AC   P42999;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ASPARTOCIN (ASPARGTOCIN).
OS   Squalus acanthias (Spiny dogfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Squaloidei; Squalidae; Squalus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 73031727.
RA   Acher R., Chauvet J., Chauvet M.-T.;
RT   "Phylogeny of the neurohypophysial hormones. Two new active peptides
RT   isolated from a cartilaginous fish, Squalus acanthias.";
RL   Eur. J. Biochem. 29:12-19(1972).
RN   [2]
RP   SEQUENCE.
RX   MEDLINE; 72128038.
RA   Acher R., Chauvet J., Chauvet M.-T., Fontaine M.;
RT   "Identification of 2 new neurohypophyseal hormones, valitocin (Val8-
RT   oxytocin) and aspartocin (Asn4-oxytocin) in a selachian fish, the
RT   spiny dog-fish (Squalus acanthias).";
RL   C. R. Acad. Sci., D, Sci. Nat. 274:313-316(1972).
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  996 MW;  17F8376EB444404B CRC64;


  Query Match             22.9%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   28.6%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        1 YLSGACL 7
            |::   |
Db        2 YINNCPL 8


RESULT 15
UXA4_CHLTR
ID   UXA4_CHLTR     STANDARD;      PRT;      5 AA.
AC   P38005;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE FROM ELEMENTARY BODY (FRAGMENT).
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE.
RC   STRAIN=L2/434/BU;
RA   Bini L., Santucci A., Magi B., Marzocchi B., Sanchez-Campillo M.,
RA   Comanducci M., Christianen G., Birkelund S., Vtretou E., Ratti G.,
RA   Pallini V.;
RL   Submitted (SEP-1994) to the SWISS-PROT data bank.
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 4.5, ITS MW IS: 28 KDA.
FT   NON_TER       5      5
SQ   SEQUENCE   5 AA;  474 MW;  75BAA865AA800000 CRC64;


  Query Match             20.8%;  Score 10;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 SG 4
            ||
Db        2 SG 3
```

```
RESULT  16
CIP2_MYTED
ID   CIP2_MYTED     STANDARD;      PRT;     6 AA.
AC   P13737;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   CONTRACTION-INHIBITING PEPTIDE II (MIP II).
OS   Mytilus edulis (Blue mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytiloidea; Mytilidae; Mytilus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PEDAL GANGLION;
RX   MEDLINE; 88240357.
RA   Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.;
RT   "Structures and actions of Mytilus inhibitory peptides.";
RL   Biochem. Biophys. Res. Commun. 152:1376-1382(1988).
CC   -!- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN
CC       MUSCLES.
CC   -!- SIMILARITY: TO MIP I.
DR   PIR; B27696; B27696.
KW   Hormone; Amidation.
FT   MOD_RES       6      6       AMIDATION.
SQ   SEQUENCE   6 AA;  621 MW;  72C9C6876DD81000 CRC64;


  Query Match             20.8%;  Score 10;  DB 1;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GA 5
           ||
Db       1 GA 2

RESULT  17
TMOF_SARBU
ID   TMOF_SARBU     STANDARD;      PRT;     6 AA.
AC   P41495;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   TRYPSIN-MODULATING OOSTATIC FACTOR (TMOF).
OS   Sarcophaga bullata (Grey flesh fly) (Neobellieria bullata).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Sarcophagidae; Sarcophaga.
RN   [1]
RP   SEQUENCE, AND SYNTHESIS.
RC   TISSUE=OVARY;
RX   MEDLINE; 94211930.
RA   Bylemans D., Borovsky D., Hunt D.F., Shabanowitz J., Grauwels L.,
RA   de Loof A.;
RT   "Sequencing and characterization of trypsin modulating oostatic
RT   factor (TMOF) from the ovaries of the grey fleshfly, Neobellieria
RT   (Sarcophaga) bullata.";
RL   Regul. Pept. 50:61-72(1994).
CC   -!- FUNCTION: HAS AN OOSTATIC ACTIVITY. INHIBITS TRYPSIN BIOSYNTHESIS
CC       IN THE MIDGUT WHICH INDIRECTLY REDUCES THE VITELLOGENIN
CC       CONCENTRATION IN THE HEMOLYMPH RESULTING IN INHIBITION OF OOCYTE
CC       DEVELOPMENT.
CC   -!- DEVELOPMENTAL STAGE: SYNTHESIZED AND RELEASED FROM FOLLICULAR
CC       EPITHELIUM AFTER A BLOOD MEAL.
KW   Hormone.
SQ   SEQUENCE   6 AA;  695 MW;  61E72451B7642000 CRC64;


  Query Match             20.8%;  Score 10;  DB 1;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       8 NL 9
           ||
Db       4 NL 5

RESULT  18
ALL2_CARMA
ID   ALL2_CARMA     STANDARD;      PRT;     7 AA.
AC   P81805;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 2.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   MOD_RES       7      7       AMIDATION (POTENTIAL).
SQ   SEQUENCE   7 AA;  770 MW;  672879CDCB5DDB70 CRC64;


  Query Match             20.8%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |  |
Db       3 YAFG 6

RESULT  19
ALL3_CARMA
ID   ALL3_CARMA     STANDARD;      PRT;     7 AA.
AC   P81806;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 3.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Multigene family.
SQ   SEQUENCE   7 AA;  796 MW;  672879CDCB476B70 CRC64;


  Query Match             20.8%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy       1 YLSG 4
           |  |
Db       3 YAFG 6


RESULT  20
ALL4_CARMA
ID   ALL4_CARMA     STANDARD;      PRT;     7 AA.
AC   P81807;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 4.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Multigene family.
SQ   SEQUENCE   7 AA;  782 MW;  672879CDCB476AC0 CRC64;


  Query Match             20.8%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |  |
Db       3 YAFG 6


RESULT  21
ALL5_CARMA
ID   ALL5_CARMA     STANDARD;      PRT;     7 AA.
AC   P81808;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 5.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   MOD_RES       7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  781 MW;  672879CDCB476420 CRC64;


  Query Match             20.8%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy       1 YLSG 4
           |  |
Db       3 YAFG 6


RESULT  22
UN06_PINPS
ID   UN06_PINPS     STANDARD;      PRT;     7 AA.
AC   P81675;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE OF NEEDLES (N141) (FRAGMENT).
OS   Pinus pinaster (Maritime pine).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Coniferopsida; Coniferales; Pinaceae; Pinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=NEEDLE;
RX   MEDLINE; 99274088.
RA   Costa P., Pionneau C., Bauw G., Dubos C., Bahrman N., Kremer A.,
RA   Frigerio J.-M., Plomion C.;
RT   "Separation and characterization of needle and xylem maritime pine
RT   proteins.";
RL   Electrophoresis 20:1098-1108(1999).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 6.6, ITS MW IS: 25 KDA.
FT   NON_TER       1      1
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  823 MW;  69D76724486B5740 CRC64;


  Query Match             20.8%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 NL 9
           ||
Db       4 NL 5


RESULT  23
AKH_MELML
ID   AKH_MELML      STANDARD;      PRT;     8 AA.
AC   P25423;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   ADIPOKINETIC HORMONE (AKH).
OS   Melolontha melolontha (Cockchafer),
OS   Geotrupes stercorosus (Dor beetle), and
OS   Pachnoda marginata (Flower beetle).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC   Scarabaeiformia; Scarabaeidae; Melolonthinae; Melolontha.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=M.MELOLONTHA, AND G.STERCOROSUS; TISSUE=CORPORA CARDIACA;
RX   MEDLINE; 91248100.
RA   Gaede G.;
RT   "A unique charged tyrosine-containing member of the adipokinetic
RT   hormone/red-pigment-concentrating hormone peptide family isolated and
RT   sequenced from two beetle species.";
RL   Biochem. J. 275:671-677(1991).
RN   [2]
RP   SEQUENCE.
RC   SPECIES=P.MARGINATA; TISSUE=CORPORA CARDIACA;
RX   MEDLINE; 92265187.
RA   Gaede G., Lopata A., Kellner R., Rinehart K.L. Jr.;
RT   "Primary structures of neuropeptides isolated from the corpora
RT   cardiaca of various cetonid beetle species determined by
RT   pulsed-liquid phase sequencing and tandem fast atom bombardment mass
```

```
RT   spectrometry.";
RL   Biol. Chem. Hoppe-Seyler 373:133-142(1992).
CC   -!- FUNCTION: THIS HORMONE, RELEASED FROM CELLS IN THE CORPORA
CC       CARDIACA AFTER THE BEGINNING OF FLIGHT, CAUSES RELEASE OF
CC       DIGLYCERIDES FROM THE FAT BODY AND THEN STIMULATES THE FLIGHT
CC       MUSCLES TO USE THESE DIGLYCERIDES AS AN ENERGY SOURCE.
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; S15422; S15422.
DR   PIR; S21663; S21663.
DR   INTERPRO; IPR002047; -.
DR   PROSITE; PS00256; AKH; 1.
KW   Neuropeptide; Amidation; Flight.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  1022 MW;  867AB775AB544736 CRC64;

  Query Match             20.8%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        7 LN 8
            ||
Db        2 LN 3


RESULT  24
AL12_CARMA
ID   AL12_CARMA     STANDARD;      PRT;     8 AA.
AC   P81815;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 12.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Multigene family.
SQ   SEQUENCE   8 AA;  913 MW;  672879CDCB569AB7 CRC64;

  Query Match             20.8%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSG 4
            |  |
Db        4 YAFG 7

RESULT  25
AL17_CARMA
ID   AL17_CARMA     STANDARD;      PRT;     8 AA.
AC   P81820;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 17.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
```

```
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   MOD_RES       8      8       AMIDATION (POTENTIAL).
SQ   SEQUENCE   8 AA;  858 MW;  C82879D5AB46D865 CRC64;

  Query Match             20.8%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 SG 4
            ||
Db        1 SG 2


RESULT  26
ALL7_CARMA
ID   ALL7_CARMA     STANDARD;      PRT;     8 AA.
AC   P81809; P81810; P81804;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 7 [CONTAINS: CARCINUSTATIN 6; CARCINUSTATIN 1].
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   CHAIN         1      8       CARCINUSTATIN 7.
FT   CHAIN         2      8       CARCINUSTATIN 6.
FT   CHAIN         4      8       CARCINUSTATIN 1.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  825 MW;  922879CDCB4775BD CRC64;

  Query Match             20.8%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSG 4
            |  |
Db        4 YAFG 7


RESULT  27
ALL8_CARMA
ID   ALL8_CARMA     STANDARD;      PRT;     8 AA.
AC   P81811;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
```

```
DE   CARCINUSTATIN 8.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  795 MW;  922879CDCB47687D CRC64;

  Query Match             20.8%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |  |
Db       4 YAFG 7


RESULT  28
ALL9_CARMA
ID   ALL9_CARMA     STANDARD;      PRT;     8 AA.
AC   P81812;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 9.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  781 MW;  7C2879CDCB476878 CRC64;

  Query Match             20.8%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |  |
Db       4 YAFG 7


RESULT  29
FAR8_CALVO
ID   FAR8_CALVO     STANDARD;      PRT;     8 AA.
AC   P41863;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 8.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; H41978; H41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  957 MW;  72D40699CAA44DD8 CRC64;

  Query Match             20.8%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GA 5
             ||
Db       1 GA 2


RESULT  30
GLUR_HUMAN
ID   GLUR_HUMAN     STANDARD;      PRT;     8 AA.
AC   P02729;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   URINE GLYCOPEPTIDE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 72062338.
RA   Lote C.J., Weiss J.B.;
RT   "Identification in urine of a low-molecular-weight highly polar
RT   glycopeptide containing cysteinyl-galactose.";
RL   Biochem. J. 123:25P-25P(1971).
CC   -!- FUNCTION: THE IDENTITY OF THE GLYCOPROTEIN FROM WHICH THIS PEPTIDE
CC       IS DERIVED IS UNKNOWN. NO PHYSIOLOGICAL FUNCTION HAS BEEN
CC       ATTRIBUTED. AN ERYTHROCYTE MEMBRANE GLYCOPEPTIDE HAVING A
CC       SIMILAR STRUCTURE HAS ALSO BEEN FOUND.
DR   PIR; A03188; XGHUEU.
KW   Glycoprotein.
FT   CARBOHYD      1      1       S-LINKED (GAL. . .).
SQ   SEQUENCE   8 AA;  855 MW;  C2D87AA1F5B1EB1E CRC64;

  Query Match             20.8%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GA 5
             ||
Db       7 GA 8


RESULT  31
HTF_TENMO
ID   HTF_TENMO      STANDARD;      PRT;     8 AA.
```

```
AC   P25419;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   HYPERTREHALOSAEMIC FACTOR (HOTH) (HYPERTREHALOSEMIC NEUROPEPTIDE).
OS   Tenebrio molitor (Yellow mealworm), and Zophobas rugipes.
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC   Cucujiformia; Tenebrionidae; Tenebrio.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CORPORA CARDIACA;
RX   MEDLINE; 90341081.
RA   Gaede G., Rosinski G.;
RT   "The primary structure of the hypertrehalosemic neuropeptide from
RT   tenebrionid beetles: a novel member of the AKH/RPCH family.";
RL   Peptides 11:455-459(1990).
CC   -!- FUNCTION: HYPERTREHALOSAEMIC FACTORS ARE NEUROPEPTIDES THAT
CC       ELEVATE THE LEVEL OF TREHALOSE IN THE HEMOLYMPH (TREHALOSE IS
CC       THE MAJOR CARBOHYDRATE IN THE HEMOLYMPH OF INSECTS).
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; A43976; A43976.
DR   PIR; B43976; B43976.
DR   INTERPRO; IPR002047; -.
DR   PROSITE; PS00256; AKH; 1.
KW   Neuropeptide; Amidation.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  1005 MW;  86745775B9C44736 CRC64;


  Query Match              20.8%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LN 8
           ||
Db       2 LN 3


RESULT  32
LCK5_LEUMA
ID   LCK5_LEUMA     STANDARD;      PRT;     8 AA.
AC   P19987;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   LEUCOKININ V (L-V).
OS   Leucophaea maderae (Madeira cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blaberoidea; Blaberidae; Leucophaea.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=HEAD;
RX   MEDLINE; 87052651.
RA   Holman G.M., Cook B.J., Nachman R.J.;
RT   "Isolation, primary structure, and synthesis of leucokinins V and VI:
RT   myotropic peptides of Leucophaea maderae.";
RL   Comp. Biochem. Physiol. 88C:27-30(1987).
CC   -!- FUNCTION: THIS CEPHALOMYOTROPIC PEPTIDE STIMULATES CONTRACTILE
CC       ACTIVITY OF COCKROACH PROTODEUM (HINDGUT).
CC   -!- SIMILARITY: TO THE OTHER LEUCOKININS.
DR   PIR; JS0315; JS0315.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  784 MW;  736365A5B9C865B8 CRC64;


  Query Match              20.8%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       3 SG 4
           ||
Db       2 SG 3


RESULT  33
LCK8_LEUMA
ID   LCK8_LEUMA     STANDARD;      PRT;     8 AA.
AC   P19990;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   LEUCOKININ VIII (L-VIII).
OS   Leucophaea maderae (Madeira cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blaberoidea; Blaberidae; Leucophaea.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=HEAD;
RA   Holman G.M., Cook B.J., Nachman R.J.;
RT   "Isolation, primary structure and synthesis of leucokinins VII and
RT   VIII: the final members of this new family of cephalomyotropic
RT   peptides isolated from head extracts of Leucophaea maderae.";
RL   Comp. Biochem. Physiol. 88C:31-34(1987).
CC   -!- FUNCTION: THIS CEPHALOMYOTROPIC PEPTIDE STIMULATES CONTRACTILE
CC       ACTIVITY OF COCKROACH PROTODEUM (HINDGUT).
CC   -!- SIMILARITY: TO THE OTHER LEUCOKININS.
DR   PIR; JS0318; JS0318.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  902 MW;  736365AB59CAADD8 CRC64;


  Query Match              20.8%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GA 5
           ||
Db       1 GA 2


RESULT  34
RPCH_PANBO
ID   RPCH_PANBO     STANDARD;      PRT;     8 AA.
AC   P08939;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RED PIGMENT CONCENTRATING HORMONE (RPCH).
OS   Pandalus borealis (Northern red shrimp).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Caridea; Pandaloidea;
OC   Pandalidae; Pandalus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 75054965.
RA   Fernlund P.;
RT   "Structure of the red-pigment-concentrating hormone of the shrimp,
RT   Pandalus borealis.";
RL   Biochim. Biophys. Acta 371:304-311(1974).
CC   -!- FUNCTION: THIS HORMONE ADAPTS THE ANIMAL TO LIGHT BACKGROUNDS BY
CC       STIMULATING CONCENTRATION OF THE PIGMENT OF ITS RED BODY-
CC       CHROMATOPHORES.
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; S07139; S07139.
DR   INTERPRO; IPR002047; -.
DR   PROSITE; PS00256; AKH; 1.
KW   Pigment; Hormone; Amidation.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  948 MW;  86786775B9C44736 CRC64;
```

```
  Query Match             20.8%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LN 8
           ||
Db       2 LN 3


RESULT  35
UF06_MOUSE
ID   UF06_MOUSE     STANDARD;      PRT;     8 AA.
AC   P38644;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE OF FIBROBLASTS (P50) (FRAGMENT).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=FIBROBLAST;
RX   MEDLINE; 95009907.
RA   Merrick B.A., Patterson R.M., Wichter L.L., He C., Selkirk J.K.;
RT   "Separation and sequencing of familiar and novel murine proteins
RT   using preparative two-dimensional gel electrophoresis.";
RL   Electrophoresis 15:735-745(1994).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 5.2, ITS MW IS: 50 KDA.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  817 MW;  A35DD878676B05B1 CRC64;


  Query Match             20.8%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GA 5
           ||
Db       6 GA 7


Search completed: December 16, 2000, 04:23:31
Job time: 4566 sec
```

```
                   GenCore version 4.5
           Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 03:35:12 ; Search time 89.11 Seconds
                                              (without alignments)
                                              6.409 Million cell updates/sec

Title:          US-09-529-121-5
Perfect score:  48
Sequence:       1 YLSGACLNL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters:      793

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       PIR_65:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                  %
Result          Query
  No.   Score   Match Length DB  ID                   Description
-------------------------------------------------------------------------
    1      17    35.4      6  2  C22565               R-phycoerythrin be
    2      17    35.4      6  2  PT0652               T-cell receptor be
    3      15    31.2      5  2  F22565               R-phycoerythrin ga
    4      15    31.2      7  2  A12016               formylglycinamide
    5      15    31.2      8  2  T13818               cytochrome oxidase
    6      15    31.2      8  2  C61512               variant surface gl
    7      15    31.2      9  2  S19329               sperm-activating p
    8      15    31.2      9  2  PT0288               Ig heavy chain CRD
    9      15    31.2      9  2  G41946               T-cell receptor ga
   10      14    29.2      5  2  B45525               actin I - malaria
   11      14    29.2      6  2  PT0605               T-cell receptor be
   12      14    29.2      6  2  PT0593               T-cell receptor be
   13      14    29.2      7  2  PT0654               T-cell receptor be
   14      14    29.2      8  2  PL0184               capsid protein VP-
   15      14    29.2      8  2  PN0043               phosphatidylethano
   16      14    29.2      9  2  C57444               neuropeptide Grb-A
   17      13    27.1      4  2  S43959               Ig mu chain V regi
   18      13    27.1      6  2  I37263               Y protein - human
   19      13    27.1      6  2  I49421               laminin B1 - weste
   20      13    27.1      7  2  S71867               glutathione transf
   21      13    27.1      7  2  S38516               mabinlin II chain
   22      13    27.1      7  2  A34026               acetylcholinestera
   23      13    27.1      7  2  I50210               gene c-rel protein
   24      13    27.1      8  2  A37521               R-phycoerythrin ga
   25      13    27.1      9  2  A61364               isotocin - common
   26      13    27.1      9  2  C41170               photosystem II pro
   27      13    27.1      9  2  A29477               diuretic neuropept
   28      13    27.1      9  4  I73804               hypothetical E2 pr
   29      12    25.0      6  2  I51434               H4 histone - Afric
   30      12    25.0      7  2  PN0649               alpha-dextrin endo
   31      12    25.0      8  2  PH1407               Ig heavy chain V r
   32      12    25.0      8  2  PQ0701               unidentified 6.5/3
   33      12    25.0      8  2  A21440               variant surface gl
   34      12    25.0      8  2  I57018               gene Cftr protein
   35      12    25.0      9  2  A44873               caldesmon - rabbit
   36      11    22.9      5  1  HOROHA               proctolin - Americ
   37      11    22.9      5  2  A41225               copper resistance
   38      11    22.9      5  2  A60411               proctolin - Atlant
   39      11    22.9      5  2  C23751               spinal cord peptid
   40      11    22.9      6  2  B44510               hypothetical prote
   41      11    22.9      6  2  PT0280               Ig heavy chain CRD
   42      11    22.9      7  2  B34818               vicilin 57K chain
   43      11    22.9      7  2  S78024               ribosomal protein
   44      11    22.9      8  2  S59622               metallothionein is
   45      11    22.9      8  2  A61467               penalbumin - Adeli
   46      11    22.9      8  2  PT0311               Ig heavy chain CRD
   47      11    22.9      8  2  B47594               aspartate kinase (
   48      11    22.9      8  2  S68325               blood cell protein
   49      11    22.9      9  2  B28495               conopressin S - co
   50      11    22.9      9  2  A43848               cell surface adhes
   51      11    22.9      9  2  S39449               retinal isomerase
   52      11    22.9      9  2  A57444               neuropeptide Grb-A
   53      11    22.9      9  2  B57444               neuropeptide Grb-A
   54      11    22.9      9  2  JN0027               [Phe-6]-mosact - s
   55      11    22.9      9  2  I50633               c-rel protein - ch
   56      11    22.9      9  2  A60427               macrophage cytotox
   57      11    22.9      9  2  PH0942               T-cell receptor be
   58      11    22.9      9  2  PH0935               T-cell receptor be
   59      11    22.9      9  2  PH0918               T-cell receptor be
   60      10    20.8      3  2  PT0571               T-cell receptor be
   61      10    20.8      4  2  PT0271               Ig heavy chain CRD
   62      10    20.8      4  2  A53284               T-cell receptor be
   63      10    20.8      4  2  PT0633               T-cell receptor be
   64      10    20.8      4  2  PT0711               T-cell receptor be
   65      10    20.8      4  2  PT0698               T-cell receptor be
   66      10    20.8      4  2  PT0677               T-cell receptor be
   67      10    20.8      4  2  PT0706               T-cell receptor be
   68      10    20.8      4  2  PT0675               T-cell receptor be
   69      10    20.8      4  2  PT0566               T-cell receptor be
   70      10    20.8      5  2  B22565               R-phycoerythrin al
   71      10    20.8      5  2  JT0520               Ig kappa chain V-I
   72      10    20.8      5  2  D44823               synaptosomal-assoc
   73      10    20.8      5  2  A26830               mitosis inhibiting
   74      10    20.8      5  2  PT0513               T-cell receptor be
   75      10    20.8      5  2  PT0608               T-cell receptor be


                               ALIGNMENTS


RESULT   1
C22565
R-phycoerythrin beta-1 chain - red alga (Gastroclonium coulteri) (fragment)
C;Species: Gastroclonium coulteri
C;Date: 07-Mar-1988 #sequence_revision 07-Mar-1988 #text_change 23-Mar-1993
C;Accession: C22565
R;Klotz, A.V.; Glazer, A.N.
J. Biol. Chem. 260, 4856-4863, 1985
A;Title: Characterization of the bilin attachment sites in R-phycoerythrin.
A;Reference number: A22565; MUID:85182601
A;Accession: C22565
A;Molecule type: protein
A;Residues: 1-6 <KLO>


  Query Match             35.4%;  Score 17;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 ACL 7
            |||
Db        3 ACL 5
```

```
RESULT   2
PT0652
T-cell receptor beta chain V-D-J region (121-1E) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0652
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0652
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-6 <FEE>
A;Experimental source: day 4 postnatal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             35.4%;  Score 17;  DB 2;  Length 6;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAC 6
           || |
Db       2 SGDC 5


RESULT   3
F22565
R-phycoerythrin gamma-A chain - red alga (Gastroclonium coulteri) (fragment)
C;Species: Gastroclonium coulteri
C;Date: 07-Mar-1988 #sequence_revision 07-Mar-1988 #text_change 23-Mar-1993
C;Accession: F22565
R;Klotz, A.V.; Glazer, A.N.
J. Biol. Chem. 260, 4856-4863, 1985
A;Title: Characterization of the bilin attachment sites in R-phycoerythrin.
A;Reference number: A22565; MUID:85182601
A;Accession: F22565
A;Molecule type: protein
A;Residues: 1-5 <KLO>


  Query Match             31.2%;  Score 15;  DB 2;  Length 5;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GAC 6
           | |
Db       1 GTC 3


RESULT   4
A12016
formylglycinamide ribonucleotide amidotransferase (EC 2.-.-.-) - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 13-Mar-1997
C;Accession: A12016; B12016
R;Ohnoki, S.; Hong, B.S.; Buchanan, J.M.
Fed. Proc. 35, 1549, 1976
A;Title: Amino acid sequence at glutamine active site for FGAR-amidotransferase.
A;Reference number: A91459
A;Accession: A12016
A;Molecule type: protein
A;Residues: 1-7 <OHN>
A;Experimental source: liver, peptide 1
A;Accession: B12016
A;Molecule type: protein
A;Residues: 1-5 <OH2>
A;Experimental source: liver, peptide 2
C;Keywords: transferase
```

```
  Query Match             31.2%;  Score 15;  DB 2;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GAC 6
           | |
Db       1 GVC 3


RESULT   5
T13818
cytochrome oxidase subunit I - Atlantic hagfish mitochondrion (fragment)
C;Species: mitochondrion Myxine glutinosa (Atlantic hagfish)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C;Accession: T13818
R;Delarbre, C.; Barriel, V.; Tillier, S.; Janvier, P.; Gachelin, G.
Mol. Biol. Evol. 14, 807-813, 1997
A;Title: The main features of the craniate mitochondrial DNA between the ND1 and the
A;Reference number: Z17775; MUID:97398704
A;Accession: T13818
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-8 <DEL>
A;Cross-references: EMBL:Y09527; NID:e1011465; PID:e329906; PIDN:CAA70718.1
C;Genetics:
A;Genome: mitochondrion
A;Note: COI
C;Keywords: mitochondrion


  Query Match             31.2%;  Score 15;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLS 3
           |||
Db       2 YLS 4


RESULT   6
C61512
variant surface glycoprotein MITat 1.4 - Trypanosoma brucei (fragment)
C;Species: Trypanosoma brucei
C;Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 07-May-1999
C;Accession: C61512
R;Holder, A.A.; Cross, G.A.M.
Mol. Biochem. Parasitol. 2, 135-150, 1981
A;Title: Glycopeptides from variant surface glycoproteins of Trypanosoma brucei. C-te
A;Reference number: A61512; MUID:81172836
A;Accession: C61512
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-8 <HOL>
C;Keywords: glycoprotein


  Query Match             31.2%;  Score 15;  DB 2;  Length 8;
  Best Local Similarity   33.3%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 SGACLN 8
           : || :
Db       3 NNACKB 8


RESULT   7
S19329
sperm-activating peptide SAP - sea urchin (Stomopneustes variolus)
C;Species: Stomopneustes variolus
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 18-Jun-1993
C;Accession: S19329
R;Yoshino, K.; Takao, T.; Shimonishi, Y.; Suzuki, N.
```

```
FEBS Lett. 294, 179-182, 1991
A;Title: Determination of the amino acid sequence of an intramolecular disulfide linkage
A;Reference number: S19329; MUID:92097763
A;Accession: S19329
A;Molecule type: protein
A;Residues: 1-9 <YOS>
F;3-8/Disulfide bonds: #status predicted


  Query Match             31.2%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 GACL 7
            | |:
Db        6 GKCV 9


RESULT   8
PT0288
Ig heavy chain CRD3 region (clone 4-106) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C;Accession: PT0288
R;Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991
A;Title: Preferential utilization of specific immunoglobulin heavy chain diversity and j
A;Reference number: PT0222; MUID:91108337
A;Accession: PT0288
A;Molecule type: DNA
A;Residues: 1-9 <YAM>
A;Experimental source: B lymphocyte
C;Keywords: heterotetramer; immunoglobulin


  Query Match             31.2%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSG 4
            | ||
Db        5 YSSG 8


RESULT   9
G41946
T-cell receptor gamma chain (2t.23) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 07-May-1999
C;Accession: G41946
R;Whetsell, M.; Mosley, R.L.; Whetsell, L.; Schaefer, F.V.; Miller, K.S.; Klein, J.R.
Mol. Cell. Biol. 11, 5902-5909, 1991
A;Title: Rearrangement and junctional-site sequence analyses of T-cell receptor gamma ge
A;Reference number: A41946; MUID:92049316
A;Accession: G41946
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-9 <WHE>
C;Keywords: T-cell receptor


  Query Match             31.2%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSG 4
            | ||
Db        5 YSSG 8


RESULT  10
B45525
actin I - malaria parasite (Plasmodium falciparum) (fragments)
```

```
C;Species: Plasmodium falciparum
C;Date: 03-Jun-1993 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
C;Accession: B45525
R;Wesseling, J.G.; Snijders, P.J.F.; van Someren, P.; Jansen, J.; Smits, M.A.; Schoen
Mol. Biochem. Parasitol. 35, 167-176, 1989
A;Title: Stage-specific expression and genomic organization of the actin genes of the
A;Reference number: A45525; MUID:89364996
A;Accession: B45525
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-5 <WES>
A;Cross-references: GB:J03988
A;Note: the authors translated the codon GAA for residue 3 as Gly
C;Comment: The actin I gene contains no introns.


  Query Match             29.2%;  Score 14;  DB 2;  Length 5;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 GAC 6
            | |
Db        2 GEC 4


RESULT  11
PT0605
T-cell receptor beta chain V-D-J region (120-1L) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0605
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0605
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-6 <FEE>
A;Experimental source: newborn thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             29.2%;  Score 14;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 SGA 5
            |||
Db        2 SGA 4


RESULT  12
PT0593
T-cell receptor beta chain V-D-J region (159-1F) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0593
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0593
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-6 <FEE>
A;Experimental source: day 19 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             29.2%;  Score 14;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       3 SGA 5
           |||
Db       4 SGA 6


RESULT  13
PT0654
T-cell receptor beta chain V-D-J region (121-1BK) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0654
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0654
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-7 <FEE>
A;Experimental source: day 4 postnatal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             29.2%;  Score 14;  DB 2;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGA 5
           |||
Db       2 SGA 4


RESULT  14
PL0184
capsid protein VP-1 - murine poliovirus (fragment)
C;Species: murine poliovirus, Theiler's encephalomyelitis virus
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Feb-1995
C;Accession: PL0184
R;Zurbriggen, A.; Hogle, J.M.; Fujinami, R.S.
J. Exp. Med. 170, 2037-2049, 1989
A;Title: Alteration of amino acid 101 within capsid protein VP-1 changes the pathogenici
A;Reference number: PL0184; MUID:90063468
A;Accession: PL0184
A;Molecule type: genomic RNA
A;Residues: 1-8 <ZUR>
C;Keywords: capsid protein


  Query Match             29.2%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 SGACLN 8
           ||   |
Db       1 SGGITN 6


RESULT  15
PN0043
phosphatidylethanol amine-binding protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 29-Oct-1997 #sequence_revision 29-Oct-1997 #text_change 23-Jan-1998
C;Accession: PN0043
R;Kato, H.
Kawasaki Igakkaishi 22, 245-259, 1996
A;Title: Analysis of proteins isolated by two dimensional electrophoresis of mouse neur
A;Reference number: PN0041
A;Accession: PN0043
A;Molecule type: protein
A;Residues: 1-8 <KAT>
A;Experimental source: neuroblastoma cell
C;Comment: The molecular mass is 23,500 and the pI is 4.8. The amino-terminus is blocke
```

```
C;Keywords: brain

  Query Match             29.2%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           |||
Db       5 LSG 7


RESULT  16
C57444
neuropeptide Grb-AST B3 - two-spotted cricket
C;Species: Gryllus bimaculatus (two-spotted cricket)
C;Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 26-Jan-1996
C;Accession: C57444
R;Lorenz, M.W.; Kellner, R.; Hoffmann, K.H.
J. Biol. Chem. 270, 21103-21108, 1995
A;Title: A family of neuropeptides that inhibit juvenile hormone biosynthesis in the
A;Reference number: A57444; MUID:95403341
A;Accession: C57444
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <LOR>


  Query Match             29.2%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           |||
Db       5 LSG 7


RESULT  17
S43959
Ig mu chain V region (clone 13) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 20-Oct-1994 #sequence_revision 26-Jul-1996 #text_change 17-Mar-1999
C;Accession: S43959
R;Wagner, S.D.; Williams, G.T.; Larson, T.; Neuberger, M.S.; Kitamura, D.; Rajewsky,
Nucleic Acids Res. 22, 1389-1393, 1994
A;Title: Antibodies generated from human immunoglobulin miniloci in transgenic mice.
A;Reference number: S43956; MUID:94248036
A;Accession: S43959
A;Molecule type: DNA
A;Residues: 1-4 <WAG>
C;Keywords: immunoglobulin


  Query Match             27.1%;  Score 13;  DB 2;  Length 4;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           | :|
Db       1 YCAG 4


RESULT  18
I37263
Y protein - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 05-Nov-1999
C;Accession: I37263
R;Waeber, G.; Habener, J.F.
Endocrinology 131, 2010-2015, 1992
A;Title: Novel testis germ cell-specific transcript of the CREB gene contains an alte
A;Reference number: I37263; MUID:93010691
A;Accession: I37263
```

```
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6 <RES>
A;Cross-references: EMBL:X68994; NID:g396171; PIDN:CAA48780.1; PID:g579816
C;Genetics:
A;Gene: CREB


  Query Match             27.1%;  Score 13;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 CL 7
           ||
Db       4 CL 5


RESULT  19
I49421
laminin B1 - western wild mouse (fragment)
C;Species: Mus spretus (western wild mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C;Accession: I49421
R;Ko, M.S.; Wang, X.; Horton, J.H.; Hagen, M.D.; Takahashi, N.; Maezaki, Y.; Nadeau, J.H
Mamm. Genome 5, 349-355, 1994
A;Title: Genetic mapping of 40 cDNA clones on the mouse genome by PCR.
A;Reference number: I48934; MUID:94319082
A;Accession: I49421
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6 <RES>
A;Cross-references: EMBL:U05736; NID:g497073; PIDN:AAB60477.1; PID:g642829


  Query Match             27.1%;  Score 13;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 CL 7
           ||
Db       5 CL 6


RESULT  20
S71867
glutathione transferase (EC 2.5.1.18) class alpha 5 - pig (fragment)
N;Alternate names: glutathione S-transferase class alpha 5
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 19-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 07-May-1999
C;Accession: S71867
R;Rouimi, P.; Anglade, P.; Debrauwer, L.; Tulliez, J.
Biochem. J. 317, 879-884, 1996
A;Title: Characterization of pig liver glutathione S-transferases using HPLC-electrospra
A;Reference number: S71864; MUID:96332484
A;Accession: S71867
A;Molecule type: protein
A;Residues: 1-7 <ROU>
C;Comment: At least five species-independent classes of cytosolic glutathion transferase
s mitochondrial form are known.
C;Complex: dimer
C;Function:
A;Description: catalyzes the nucleophilic conjugation of intracellular glutathione to a
A;Pathway: detoxification; xenobiotics metabolism
A;Note: increased hydrophilicity of GSH-conjugates facilitates their further metabolism
es of damage
C;Superfamily: glutathione transferase
C;Keywords: dimer; transferase


  Query Match             27.1%;  Score 13;  DB 2;  Length 7;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy       5 ACLN 8
           | ||
Db       1 AILN 4


RESULT  21
S38516
mabinlin II chain A - Yunnan caper (fragments)
C;Species: Capparis masaikai (Yunnan caper)
C;Date: 08-Jun-1994 #sequence_revision 27-Feb-1997 #text_change 13-Mar-1997
C;Accession: S38516
R;Nirasawa, S.; Liu, X.; Nishino, T.; Kurihara, Y.
Biochim. Biophys. Acta 1202, 277-280, 1993
A;Title: Disulfide bridge structure of the heat-stable sweet protein mabinlin II.
A;Reference number: S38516; MUID:94002261
A;Accession: S38516
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-7 <NIR>


  Query Match             27.1%;  Score 13;  DB 2;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 AC 6
           ||
Db       4 AC 5


RESULT  22
A34026
acetylcholinesterase (EC 3.1.1.7) 5.6S form - Pacific electric ray (fragment)
C;Species: Torpedo californica (Pacific electric ray)
C;Date: 16-Mar-1990 #sequence_revision 16-Mar-1990 #text_change 08-Nov-1996
C;Accession: A34026
R;Gibney, G.; MacPhee-Quigley, K.; Thompson, B.; Vedvick, T.; Low, M.G.; Taylor, S.S.
J. Biol. Chem. 263, 1140-1145, 1988
A;Title: Divergence in primary structure between the molecular forms of acetylcholine
A;Reference number: A34026; MUID:88087239
A;Accession: A34026
A;Molecule type: protein
A;Residues: 1-7 <GIB>
C;Keywords: alternative splicing; carboxylic ester hydrolase


  Query Match             27.1%;  Score 13;  DB 2;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 AC 6
           ||
Db       6 AC 7


RESULT  23
I50210
gene c-rel protein - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 28-Feb-1997
C;Accession: I50210
R;Kabrun, N.; Bumstead, N.; Hayman, M.J.; Enrietto, P.J.
Mol. Cell. Biol. 10, 4788-4794, 1990
A;Title: Characterization of a novel promoter insertion in the c-rel locus.
A;Reference number: I50210; MUID:90355995
A;Accession: I50210
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-7 <KAB>
A;Cross-references: GB:M55577; NID:g555438; PID:g211661
C;Genetics:
A;Gene: c-rel
```

```
  Query Match             27.1%;  Score 13;  DB 2;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGA 5
           ::||
Db       1 MAGA 4


RESULT  24
A37521
R-phycoerythrin gamma-E chain - red alga (Gastroclonium coulteri) (fragment)
C;Species: Gastroclonium coulteri
C;Date: 07-Mar-1988 #sequence_revision 07-Mar-1988 #text_change 23-Mar-1993
C;Accession: A37521; J22565
R;Klotz, A.V.; Glazer, A.N.
J. Biol. Chem. 260, 4856-4863, 1985
A;Title: Characterization of the bilin attachment sites in R-phycoerythrin.
A;Reference number: A22565; MUID:85182601
A;Accession: A37521
A;Molecule type: protein
A;Residues: 1-8 <KLO>

  Query Match             27.1%;  Score 13;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 AC 6
           ||
Db       6 AC 7


RESULT  25
A61364
isotocin - common carp
C;Species: Cyprinus carpio (common carp)
C;Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 20-Jun-2000
C;Accession: A61364
R;Acher, R.; Chauvet, J.; Chauvet, M.T.; Crepy, D.
Comp. Biochem. Physiol. A 14, 245-254, 1965
A;Title: Caracterisation des hormones neurohypophysaires d'un poisson osseux d'eau douce
A;Reference number: A61364
A;Accession: A61364
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <ACH>
C;Superfamily: oxytocin-neurophysin
C;Keywords: amidated carboxyl end; neuropeptide; posterior pituitary
F;9/Modified site: amidated carboxyl end (Gly) #status experimental

  Query Match             27.1%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLS 3
           |:|
Db       2 YIS 4

RESULT  26
C41170
photosystem II protein psbM - Chlamydomonas reinhardtii (fragment)
C;Species: Chlamydomonas reinhardtii
C;Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 30-Sep-1993
C;Accession: C41170
R;de Vitry, C.; Diner, B.A.; Popot, J.L.
J. Biol. Chem. 266, 16614-16621, 1991
A;Title: Photosystem II particles from Chlamydomonas reinhardtii. Purification, molecula
A;Reference number: A41170; MUID:91358452
A;Accession: C41170
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <DE5>

  Query Match             27.1%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGA 5
           ::||
Db       4 IAGA 7


RESULT  27
A29477
diuretic neuropeptide F1 - migratory locust
C;Species: Locusta migratoria (migratory locust)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 31-Dec-1993
C;Accession: A29477
R;Proux, J.P.; Miller, C.A.; Li, J.P.; Carney, R.L.; Girardie, A.; Delaage, M.; Schoo
Biochem. Biophys. Res. Commun. 149, 180-186, 1987
A;Title: Identification of an arginine vasopressin-like diuretic hormone from Locusta
A;Reference number: A29477; MUID:88077077
A;Accession: A29477
A;Molecule type: protein
A;Residues: 1-9 <PRO>
A;Note: two neuropeptides, F1 and F2, were identified. F2 is an antiparallel dimer of
C;Keywords: neuropeptide


  Query Match             27.1%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 CL 7
           ||
Db       1 CL 2


RESULT  28
I73804
hypothetical E2 protein - human papillomavirus type 16 (fragment)
C;Species: human papillomavirus type 16
C;Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 18-Feb-2000
C;Accession: I73804
R;Schneider-Maunoury, S.; Croissant, O.; Orth, G.
J. Virol. 61, 3295-3298, 1987
A;Title: Integration of human papillomavirus type 16 DNA sequences: a possible early
A;Reference number: I56695; MUID:87311896
A;Accession: I73804
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-9 <SCH>
A;Cross-references: GB:M31225; NID:g190254; PIDN:AAA65996.1; PID:g553617
C;Comment: This is the hypothetical translation of a viral sequence integrated into t


  Query Match             27.1%;  Score 13;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 CL 7
           ||
Db       7 CL 8

RESULT  29
I51434
H4 histone - African clawed frog (fragment)
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Sep-1996
C;Accession: I51434
```

```
R;Woodland, H.R.; Warmington, J.R.; Ballantine, J.E.M.; Turner, P.C.
Nucleic Acids Res. 12, 4939-4958, 1984
A;Title: Are there major developmentally regulated H4 gene classes in Xenopus?.
A;Reference number: I51391; MUID:84247348
A;Accession: I51434
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-6 <WOO>
A;Cross-references: GB:K02304; NID:g214227; PID:g555517


  Query Match             25.0%;  Score 12;  DB 2;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       1 MSG 3


RESULT  30
PN0649
alpha-dextrin endo-1,6-alpha-glucosidase (EC 3.2.1.41) - Bacillus sp. (strain S-1) (frag
C;Species: Bacillus sp.
C;Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 07-May-1999
C;Accession: PN0649
R;Kim, C.H.; Choi, H.I.; Lee, D.S.
Biosci. Biotechnol. Biochem. 57, 1632-1637, 1993
A;Title: Purification and biochemical properties of an alkaline pullulanase from alkalop
A;Reference number: PN0649; MUID:94080025
A;Accession: PN0649
A;Molecule type: protein
A;Residues: 1-7 <KIM>
C;Comment: This enzyme is used togther with glucoamylase to improve the efficiency of st
ntent in high maltose syrups.
C;Keywords: glycosidase; hydrolase; polysaccharide degradation


  Query Match             25.0%;  Score 12;  DB 2;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LNL 9
           ||:
Db       2 LNM 4


RESULT  31
PH1407
Ig heavy chain V region (clone micro m+ 46-6 PCR) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 17-Mar-1999
C;Accession: PH1407
R;Shirasawa, T.; Miyazoe, I.; Hagiwara, S.; Kimoto, H.; Shigemoto, K.; Taniguchi, M.; Ta
J. Exp. Med. 176, 1209-1214, 1992
A;Title: Heavy chain variable (VH) region diversity generated by VH gene replacement in
ia virus.
A;Reference number: PH1403; MUID:93018837
A;Accession: PH1407
A;Molecule type: DNA
A;Residues: 1-8 <SHI>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin


  Query Match             25.0%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 CLNL 9
           |  |
Db       2 CARL 5
```

```
RESULT  32
PQ0701
unidentified 6.5/31K protein [imported] - rice (fragment)
C;Species: Oryza sativa (rice)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: PQ0701
R;Komatsu, S.; Kajiwara, H.; Hirano, H.
Theor. Appl. Genet. 86, 935-942, 1993
A;Title: A rice protein library; a data-file of rice proteins separated by two-dimens
A;Reference number: PQ0696
A;Accession: PQ0701
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-8 <KOM>


  Query Match             25.0%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGA 5
           ::||
Db       5 VTGA 8


RESULT  33
A21440
variant surface glycoprotein pSLc1 - Trypanosoma brucei (fragment)
C;Species: Trypanosoma brucei
C;Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 20-Mar-1998
C;Accession: A21440
R;Parsons, M.; Nelson, R.G.; Watkins, K.P.; Agabian, N.
Cell 38, 309-316, 1984
A;Title: Trypanosome mRNAs share a common 5' spliced leader sequence.
A;Reference number: A90853; MUID:84282716
A;Accession: A21440
A;Molecule type: mRNA
A;Residues: 1-8 <PAR>
A;Cross-references: GB:K02195; NID:g162150; PID:g162151
C;Keywords: glycoprotein


  Query Match             25.0%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       1 MSG 3


RESULT  34
I57018
gene Cftr protein - mouse (fragment)
C;Species: Mus sp. (mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 28-Feb-1997
C;Accession: I57018
R;Dorin, J.R.; Stevenson, B.J.; Fleming, S.; Alton, E.W.; Dickinson, P.; Porteous, D.
Mamm. Genome 5, 465-472, 1994
A;Title: Long-term survival of the exon 10 insertional cystic fibrosis mutant mouse i
A;Reference number: I57018; MUID:95037043
A;Accession: I57018
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-8 <RES>
A;Cross-references: GB:S74246; NID:g710482
C;Genetics:
A;Gene: Cftr


  Query Match             25.0%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
```

```
  Matches     2;  Conservative     1;  Mismatches     1;  Indels     0;  Gaps     0;

Qy        3 SGAC 6
            | :|
Db        4 SPSC 7


RESULT  35
A44873
caldesmon - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 31-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 31-Oct-1997
C;Accession: A44873
R;Ikebe, M.; Hornick, T.
Arch. Biochem. Biophys. 288, 538-542, 1991
A;Title: Determination of the phosphorylation sites of smooth muscle caldesmon by protei
A;Reference number: A44873; MUID:91378498
A;Accession: A44873
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <IKE>
A;Experimental source: skeletal myosin
A;Note: sequence extracted from NCBI backbone (NCBIP:63199)
C;Superfamily: caldesmon


  Query Match             25.0%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   33.3%;  Pred. No. 1.8e+05;
  Matches     2;  Conservative     2;  Mismatches     2;  Indels     0;  Gaps     0;

Qy        4 GACLNL 9
            |: | :
Db        1 GSSLKI 6


Search completed: December 16, 2000, 03:35:13
Job time: 5646 sec
```

```
                    GenCore version 4.5
           Copyright (c) 1993 - 2000  Compugen Ltd.


OM protein - protein search, using sw model

Run on:         December 16, 2000, 03:07:19 ; Search time 107.12 Seconds
                                              (without alignments)
                                              2.873 Million cell updates/sec

Title:          US-09-529-121-5
Perfect score:  48
Sequence:       1 YLSGACLNL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters:      61695

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       A_Geneseq_36:*
                1:  /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
                2:  /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
                3:  /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
                4:  /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
                5:  /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
                6:  /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
                7:  /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
                8:  /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
                9:  /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
               10:  /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
               11:  /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
               12:  /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
               13:  /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
               14:  /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
               15:  /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
               16:  /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
               17:  /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
               18:  /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
               19:  /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
               20:  /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
               21:  /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                %
Result          Query
  No.   Score   Match Length DB  ID                   Description
-------------------------------------------------------------------------
   1       48   100.0      9  20  Y09529              Carcinoembryonic a
   2       36    75.0      9  18  W39723              Human carcina-embr
   3       36    75.0      9  19  W77134              CEA synthetic pept
   4       36    75.0      9  19  W70045              CEA derived HLA-A2
   5       36    75.0      9  20  Y47655              Immunogenic peptid
   6       36    75.0      9  20  Y09525              Carcinoembryonic a
   7       36    75.0      9  20  Y09526              Carcinoembryonic a
   8       34    70.8      9  20  Y09527              Carcinoembryonic a
   9       34    70.8      9  20  Y09528              Carcinoembryonic a
  10       33    68.8      9  21  Y54173              HLA binding peptid
  11       29    60.4      9  17  W00680              Peptide comprising
  12       26    54.2      7  19  W76240              Bacterial periplas
  13       26    54.2      9  21  Y82882              Teratocarcinoma-de
  14       24    50.0      7  20  Y49030              Membrane dipeptida
  15       24    50.0      7  20  Y39584              CTLA-4 VLD CDR1 re
  16       24    50.0      9  19  W70043              MAGE 3 antigen der
  17       24    50.0      9  20  Y47153              Immunogenic peptid
  18       24    50.0      9  20  Y47480              Immunogenic peptid
  19       24    50.0      9  20  Y47562              Immunogenic peptid
  20       24    50.0      9  20  Y47563              Immunogenic peptid
  21       23    47.9      6  17  Y01944              Fragment of the EG
  22       23    47.9      8  14  R37279              PM-1/BSA peptide.
  23       23    47.9      8  14  R37280              PM-1/BSA peptide.
  24       23    47.9      8  14  R37281              PM-1/BSA peptide.
  25       23    47.9      8  14  R37282              PM-1/BSA peptide.
  26       23    47.9      8  14  R37283              PM-1/BSA peptide.
  27       23    47.9      8  14  R37284              PM-1/BSA peptide.
  28       23    47.9      8  14  R37285              PM-1/BSA peptide.
  29       23    47.9      8  14  R37286              PM-1/BSA peptide.
  30       23    47.9      8  14  R37287              PM-1/BSA peptide.
  31       23    47.9      8  18  W13449              Kidney homing pept
  32       23    47.9      8  19  W77008              Peptide mimetic of
  33       23    47.9      9  19  W40267              K. oxytoca R-speci
  34       23    47.9      9  21  Y82878              Teratocarcinoma-de
  35       22    45.8      9  17  W49470              Human leucocyte an
  36       22    45.8      9  20  Y47757              Immunogenic peptid
  37       21    43.8      6  11  R09414              LFA-1 alpha subuni
  38       21    43.8      6  17  R88476              Internal tryptic p
  39       21    43.8      7  17  W04985              Immunomodulatory p
  40       21    43.8      7  18  W44992              Immunomodulatory p
  41       21    43.8      7  18  W12016              Herpes virus DNA p
  42       21    43.8      7  20  Y09418              Immunoactive pepti
  43       21    43.8      8  14  R36235              Hepatitis C virus
  44       21    43.8      8  18  W30073              GDNF, neurturin, p
  45       21    43.8      8  20  Y16679              WO9914235 Seq ID N
  46       21    43.8      8  20  Y16687              WO9914235 Seq ID N
  47       21    43.8      9  15  R61804              PLP peptide 161, p
  48       21    43.8      9  16  R87430              Human MHC class II
  49       21    43.8      9  16  R80907              MAGE-2 peptide (re
  50       21    43.8      9  17  W04992              Immunomodulatory p
  51       21    43.8      9  18  W44997              Immunomodulatory p
  52       21    43.8      9  18  W36558              MAGE-2 peptide (re
  53       21    43.8      9  20  Y48863              Membrane dipeptida
  54       21    43.8      9  20  Y47313              Immunogenic peptid
  55       21    43.8      9  20  Y16690              WO9914235 Seq ID N
  56       21    43.8      9  20  Y09423              Immunoactive pepti
  57       21    43.8      9  20  W97825              Plant expansin pep
  58       20    41.7      7  17  W04984              Immunomodulatory p
  59       20    41.7      7  18  W44991              Immunomodulatory p
  60       20    41.7      7  18  W13451              Kidney homing pept
  61       20    41.7      7  20  Y09417              Immunoactive pepti
  62       20    41.7      7  21  Y78206              Modulating agent c
  63       20    41.7      9  15  R61827              PLP peptide 107, p
  64       20    41.7      9  16  R87433              Human MHC class II
  65       20    41.7      9  16  R79851              Test peptide #4 to
  66       20    41.7      9  16  R77581              HIV-B51-71(H-51),
  67       20    41.7      9  16  R70563              HIV peptide fragme
  68       20    41.7      9  17  W49644              Human leucocyte an
  69       20    41.7      9  17  W00690              NCA analogue of re
  70       20    41.7      9  20  Y53336              Bcr-Abl epitope (a
  71       20    41.7      9  20  Y47152              Immunogenic peptid
  72       20    41.7      9  20  Y47422              Immunogenic peptid
  73       20    41.7      9  20  Y47532              Immunogenic peptid
  74       20    41.7      9  20  Y40158              Amino acid sequenc
  75       20    41.7      9  20  Y26674              BCR-ABL-derived li

                              ALIGNMENTS


RESULT   1
Y09529
ID   Y09529 standard; peptide; 9 AA.
XX
AC   Y09529;
```

```
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:5.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match            100.0%;  Score 48;  DB 20;  Length 9;
  Best Local Similarity  100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGACLNL 9
            |||||||||
Db        1 ylsgaclnl 9

RESULT   2
W39723
ID   W39723 standard; peptide; 9 AA.
XX
AC   W39723;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Human carcina-embryonic antigen (CEA) peptide (pos. 571-579).
XX
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
XX
OS   Homo sapiens.
XX
```

```
PN   WO9741440-A1.
XX
PD   06-NOV-1997.
XX
PF   28-APR-1997;   97WO-NL00229.
XX
PR   23-DEC-1996;   96EP-0203670.
PR   26-APR-1996;   96EP-0201145.
XX
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
PI   Kast WM,  Melief CJM,  Offringa R,  Toes REM,  Van Der Burg SH;
XX
DR   WPI; 1997-549891/50.
XX
PT   Method of selecting T cell peptide epitope(s) - by measuring the
PT   stability of HLA class I-peptide complexes on intact B cells
XX
PS   Example 3; Page 85; 109pp; English.
XX
CC   Peptides W39430-W39734 are used in a novel method for the selection of
CC   immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC   method involves the identification of peptide sequences capable of
CC   binding to an HLA (human leukocyte antigen) class I molecule and
CC   measuring the binding of this epitope peptide to the HLA class I peptide.
CC   The stability of binding of the peptide and MHC (major histocompatablilty
CC   complex) class I molecule is measured on intact human B cells carrying
CC   the MHC molecule at their cell surfaces. The method can be used to select
CC   peptide epitopes for generating vaccines against a disease associated
CC   with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC   especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC   immune responses. Peptide W39723 is derived from the human
CC   carcino-embryonic antigen (CEA) and has the ability to bind to the human
CC   MHC Class I allele HLA-A2.1.
XX
SQ   Sequence   9 AA;

  Query Match            75.0%;  Score 36;  DB 18;  Length 9;
  Best Local Similarity  88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGACLNL 9
            ||||| |||
Db        1 ylsganlnl 9

RESULT   3
W77134
ID   W77134 standard; peptide; 9 AA.
XX
AC   W77134;
XX
DT   16-NOV-1998  (first entry)
XX
DE   CEA synthetic peptide epitope 1.
XX
KW   Tyrosinase; tyrosinase cytotoxic lymphocyte response;
KW   cytotoxic T lymphocyte; cysteine-depleted; melanoma.
XX
OS   Synthetic.
XX
PN   WO9833810-A2.
XX
PD   06-AUG-1998.
XX
PF   29-JAN-1998;   98WO-US01592.
XX
PR   30-JAN-1997;   97US-0037781.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
```

```
PI   Engelhard VH,  Hunt DF,  Kittlesen D,  Slingluff CL;
XX
DR   WPI; 1998-437388/37.
XX
PT   Disease specific immunogen  - comprises disease specific cytotoxic T
PT   lymphocyte epitope used to elicit melanoma specific CTL response
XX
PS   Disclosure; Page 27; 93pp; English.
XX
CC   The peptide epitope W77119-W77138 were created for human tumour-specific
CC   cytotoxic T lymphocyte response.  These peptides are are cysteine-
CC   depleted mutants of a native disease-specific CTL epitope.  The cysteine-
CC   depleted CTL epitopes elicit a stronger or more specific CTL response
CC   than the native epitope.  The epitopes can be used in a disease-specific
CC   immunogen to protect a mammal against disease in particular melanomas.
CC   The peptides may also be used to screen a sample for the presence of
CC   an antigen with the same epitope, or with a different cross-reactive
CC   epitope.
XX
SQ   Sequence   9 AA;


  Query Match             75.0%;  Score 36;  DB 19;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
                |||||  |||
Db          1 ylsganlnl 9


RESULT   4
W70045
ID   W70045 standard; peptide; 9 AA.
XX
AC   W70045;
XX
DT   22-OCT-1998  (first entry)
XX
DE   CEA derived HLA-A2.1 binding peptide 2 (residues 605-613).
XX
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW   antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW   viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW   fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9833888-A1.
XX
PD   06-AUG-1998.
XX
PF   30-JAN-1998;   98WO-US01959.
XX
PR   31-JAN-1997;   97US-0036696.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Celis E,  Sette A,  Sidney J,  Southwood S,  Tsai V;
XX
DR   WPI; 1998-437445/37.
XX
PT   Production of antigen-specific cytotoxic T cells - by incubating
PT   immunogenic peptide(s) from antigen that binds class I major
PT   histocompatibility complex molecules with pre-treated antigen
PT   presenting cells
XX
PS   Example 6; Page 75; 104pp; English.
XX
CC   Sequences shown in W70044 to W70052 represent peptides derived from
CC   carcinoembryonic antigen (CEA). The peptides can bind to a human
CC   leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method
CC   of invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC   vitro. The method comprises contacting immunogenic peptides from an
CC   antigen that binds class I major histocompatibility complex (MHC)
CC   molecules with antigen presenting cells (APCs) pretreated with
CC   pretreatment growth factors, and incubating the APCs with purified CD8
CC   cells in the presence of at least 2 incubation growth factors, thereby
CC   producing antigen-specific CTLs. A method for specifically killing
CC   target cells in a human patient is also provided which comprises
CC   obtaining a fluid sample containing CTLs from a patient, contacting the
CC   cytotoxic T cells with APCs pretreated with pre-treatment growth factors,
CC   where the APCs comprise class I MHC molecules. The pretreated APCs are
CC   incubated with the cytotoxic growth factors, thereby producing activated
CC   CTLs which are contacted with a carrier to form a composition. The
CC   composition can then be administered to the patient. The activated CTLs
CC   can be used for treating cancers, immune disorders, viral infections,
CC   AIDS, hepatitis, bacterial infection, fungal infection, malaria or
CC   tuberculosis.
XX
SQ   Sequence   9 AA;


  Query Match             75.0%;  Score 36;  DB 19;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGACLNL 9
                |||||  |||
Db          1 ylsganlnl 9


RESULT   5
Y47655
ID   Y47655 standard; Peptide; 9 AA.
XX
AC   Y47655;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2266.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 118; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
```

```
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;

  Query Match             75.0%;  Score 36;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
           ||||| |||
Db       1 ylsganlnl 9

RESULT   6
Y09525
ID   Y09525 standard; peptide; 9 AA.
XX
AC   Y09525;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist CAP-1.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 1; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
```

```
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             75.0%;  Score 36;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
           ||||| |||
Db       1 ylsganlnl 9

RESULT   7
Y09526
ID   Y09526 standard; peptide; 9 AA.
XX
AC   Y09526;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:2.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             75.0%;  Score 36;  DB 20;  Length 9;
```

```
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
           ||||| |||
Db       1 ylsgadlnl 9


RESULT   8
Y09527
ID   Y09527 standard; peptide; 9 AA.
XX
AC   Y09527;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:3.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             70.8%;  Score 34;  DB 20;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
           ||||| :||
Db       1 ylsgadinl 9


RESULT   9
Y09528
ID   Y09528 standard; peptide; 9 AA.
XX
AC   Y09528;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:4.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             70.8%;  Score 34;  DB 20;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
           ||||| :||
Db       1 ylsganinl 9


RESULT  10
Y54173
ID   Y54173 standard; peptide; 9 AA.
XX
AC   Y54173;
XX
DT   06-APR-2000  (first entry)
XX
DE   HLA binding peptide 1233.11 derived from source CEA.605V9.
XX
KW   Allele-specific binding motif; major histocompatibility complex; MHC;
KW   HLA; HLA-A2.1; cytotoxic T cell response; antigen; microbial infection;
KW   hepatitis; AIDS; malaria; condyloma acuminatum; cancer; prostate cancer;
KW   renal carcinoma; cervical carcinoma; lymphoma; tumour.
```

```
XX
OS   Unidentified.
XX
PN   WO9965522-A1.
XX
PD   23-DEC-1999.
XX
PF   17-JUN-1999;   99WO-US13789.
XX
PR   17-JUN-1998;   98US-0098584.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Sidney J,  Southwood S;
XX
DR   WPI; 2000-106018/09.
XX
PT   Novel HLA binding immunogenic peptides used to induce T cell activation
PT   and to induce an immune response  -
XX
PS   Claim 1; Page 32; 42pp; English.
XX
CC   Peptides Y54171-Y54236 represent immunogenic peptides comprising an
CC   allele-specific binding motif for the major histocompatibility complex
CC   (MHC) molecule HLA, e.g. HLA-A2.1. The peptides have conserved residues
CC   at certain positions such as positions 2 and 9. Also, the peptides do not
CC   comprise negative binding residues at other positions, such as positions
CC   1, 3, 6 and/or 7 (peptides 9 amino acids long) and at poisons 1, 3, 4,
CC   5, 7, 8 and/or 9 (peptides 10 amino acids long). The peptides are used to
CC   induce a cytotoxic T cell response to a preselected antigen. The method
CC   comprises contacting cytotoxic T cells from a patient (optionally
CC   expressing a specific MHC class I allele) with the present peptides.
CC   The peptides are used to treat and prevent microbial infection (e.g. in
CC   viral hepatitis B and C, human papilomavirus (HPV) infection, AIDS,
CC   cytomegalovirus (CMV), malaria, and condyloma acuminatum) and cancer
CC   (e.g. prostate cancer, renal carcinoma, cervical carcinoma, lymphoma).
CC   Patients in the acute phase of infection can be treated with the
CC   peptides in conjunction with other treatments. The antigenic peptides
CC   may be used to elicit cytotoxic T lymphocytes (CTLs) ex vivo and in
CC   vivo. The resulting CTLs can be used to treat chronic infections (viral
CC   or bacterial) or tumours in patients that do not respond to conventional
CC   forms of therapy. The peptides may also be used to produce monoclonal
CC   antibodies, which are useful as potential diagnostic or therapeutic
CC   agents. The peptides may also be used as diagnostic reagents.
XX
SQ   Sequence   9 AA;


  Query Match             68.8%;  Score 33;  DB 21;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         1 YLSGACLNL 9
             ||||| ||:
Db         1 ylsganlnv 9


RESULT  11
W00680
ID   W00680 standard; peptide; 9 AA.
XX
AC   W00680;
XX
DT   01-MAY-1997  (first entry)
XX
DE   Peptide comprising residues 571-579 of Carcinoembryonic antigen.
XX
KW   Carcinoembryonic; antigen; human; cytotoxic T cell; pox virus;
KW   vector; epitope; determination; screening; tumour; treatment.
XX
OS   Homo sapiens.
XX
PN   WO9626271-A1.
```

```
XX
PD   29-AUG-1996.
XX
PF   13-FEB-1996;   96WO-US02156.
XX
PR   22-FEB-1995;   95US-0396385.
XX
PA   (THER-) THERION BIOLOGICS CORP.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Panicali D,  Schlom J,  Tsang KY;
XX
DR   WPI; 1996-402364/40.
XX
PT   Generation of human cytotoxic T-cells specific for CEA - useful in
PT   therapy, epitope mapping and drug screening
XX
PS   Claim 4; Page 57; 76pp; English.
XX
CC   Producing carcinoembryonic antigen (CEA) specific human cytotoxic T
CC   cells (CTC), comprises introducing a 1st pox virus vector, having
CC   at least 1 insertion site containing a DNA segment encoding a CEA
CC   peptide (i.e. the present peptide) to a host to stimulate CTC
CC   production, and at least 1 periodic interval after that, contacting
CC   the host with an additional antigen. The CEA specific CTC can be
CC   used to determine the CTC eliciting epitope of CEA, and to screen
CC   for compounds which enhance the ability of the antigen to create a
CC   CTC response. A host with a CEA expressing tumour can be treated by
CC   introducing the CTC to the host, and at least 1 periodic interval
CC   after that introducing a CEA peptide, i.e. the present peptide.
CC   The present peptide is positive for binding to HLA-A2, and scored
CC   561 and 806 in T2 cell binding assays, where the binding of an
CC   appropriate peptide results in the upregulation of surface HLA-A2
CC   on the T2 cells, which can be quantified via FACScan using an
CC   anti-HLA-A2 antibody (background 280 and 300).
XX
SQ   Sequence   9 AA;


  Query Match             60.4%;  Score 29;  DB 17;  Length 9;
  Best Local Similarity   87.5%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         2 LSGACLNL 9
             |||| |||
Db         2 lsganlnl 9


RESULT  12
W76240
ID   W76240 standard; protein; 7 AA.
XX
AC   W76240;
XX
DT   02-DEC-1998  (first entry)
XX
DE   Bacterial periplasmic binding protein fragment #15.
XX
KW   Protein-ligand binding pocket; PLBP; binding protein; ligand; modulator;
KW   bacterial periplasmic binding protein; interaction energies; ischaemia;
KW   basis set molecules; BSM; ionotropic glutamate receptors; treatment;
KW   neuroprotectant; stroke; epilepsy; neuropathic pain;
XX
OS   Prokaryota.
XX
PN   WO9838208-A2.
XX
PD   03-SEP-1998.
XX
PF   27-FEB-1998;   98WO-US03951.
XX
PR   28-FEB-1997;   97US-0808804.
XX
```

```
PA   (BEAR-) BEARSDEN BIO INC.
XX
PI   Sturgess M;
XX
DR   WPI; 1998-495386/42.
XX
PT   Designing a model for a ligand-binding pocket in a protein and its
PT   use for assessing ligand affinity - without the need to prepare test
PT   ligand, for identifying selective antagonists for ionotropic
PT   glutamate receptors, potentially useful as neuroprotectants with
PT   reduced side effects
XX
PS   Disclosure; Page 216; 218pp; English.
XX
CC   W76226-W76250 are protein fragments used in the design of a model of a
CC   protein-ligand binding pocket (PLBP) of a binding protein. A model is
CC   constructed based on typographic similarity to the binding pocket of a
CC   bacterial periplasmic binding protein. The model is refined by energy
CC   minimisation, with a high affinity ligand in the binding pocket.
CC   Interaction energies of basis set molecules (BSM) are obtained by
CC   calculating energy of the models for the binding pocket BP and BSM
CC   individually (Er and Em) and total energy (Erm) of bound complexes
CC   formed. The model is used to predict the selectivity of a potential
CC   ligand for a set of related PLBP's, specifically ionotropic glutamate
CC   receptors. Identified ligands are potentially useful for studying
CC   receptor binding and activity and as modulators of receptor activity
CC   (or lead compounds for developing such compounds). The ligands are
CC   potentially useful therapeutically, e.g. as neuroprotectants during
CC   ischaemia and for treatment of stroke, head injuries, epilepsy,
CC   neuropathic pain etc. The method allows affinity of ligands to be
CC   estimated without having to prepare them and then test them, in vitro.
CC   Ligands with high selectivity for particular glutamate receptors
CC   should have fewer side effects than known receptor antagonists.
XX
SQ   Sequence   7 AA;


  Query Match             54.2%;  Score 26;  DB 19;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 GACLNL 9
          |||::|
Db      1 gacidl 6


RESULT  13
Y82882
ID   Y82882 standard; peptide; 9 AA.
XX
AC   Y82882;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Teratocarcinoma-derived growth factor (CRIPTO-1) antigenic peptide.
XX
KW   Tumour associated antigen peptide; TAA; cancer; carcinoma;
KW   treatment; prevention; cure; anti-tumour vaccine; metastases;
KW   breast; bladder; prostate; pancreas; ovary; thyroid; colon;
KW   stomach; carcinoma; MHC Class I; HLA-A2; human;
KW   Major Histocompatibility Complex; uroplakin;
KW   prostate specific antigen; prostate specific membrane antigen;
KW   prostate acid phosphatase; mucin; lactadherin;
KW   teratocarcinoma derived growth factor; PSA; PSMA; PAP; CRIPTO-1.
XX
OS   Homo sapiens.
XX
PN   WO200006723-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-IL00417.
XX
```

```
PR   30-JUL-1998;   98IL-0125608.
XX
PA   (YEDA ) YEDA RES & DEV CO LTD.
PA   (BIOT-) BIO-TECHNOLOGY GEN CORP.
XX
PI   Eisenbach L,  Carmon L,  Tirosh B,  Bar-haim E,  Paz A,  Fridkin M;
PI   Fitzer-attas C;
XX
DR   WPI; 2000-205463/18.
XX
PT   Tumor associated antigen peptides, especially derived from uroplakin,
PT   useful as vaccines to prevent or cure cancers including breast,
PT   bladder, prostate, pancreas, ovary, thyroid, colon and stomach -
XX
PS   Claim 19; Page 108; 113pp; English.
XX
CC   Tumour associated antigen peptides (TAA) may be used for the
CC   treatment, prevention and cure of cancer or cancer metastases. The
CC   cancer may be breast, bladder, prostate, pancreas, ovary, thyroid,
CC   colon, stomach, head or neck cancer or a carcinoma. The tumour
CC   associated antigens are presentable to the immune system by HLA-A2
CC   molecules and are generally between 8 to 10 amino acids in length.
CC   The amino acids located at positions 2 and 9 of the tumour associated
CC   antigens are the anchor residues which participate in the binding to
CC   MHC class I molecules, more specifically HLA-A2.  More tumour
CC   associated antigens are described in GENESEQ records Y82806-Y82882.
CC   Those tumour associated antigens described in records Y82806-Y82824
CC   and Y82855-YY82869 are derived from Uroplakin, such as Uroplakin II,
CC   Uroplakin Ia, Uroplakin III and Uroplakin Ib.  Those described in
CC   records Y82825-Y82829 are derived from prostate specific antigen
CC   (PSA). Those described in records Y82830-Y82835 are derived from
CC   prostate specific membrane antigen (PSMA).  Those described in
CC   records YY82836-Y82839 are derived from prostate acid phosphotase
CC   (PAP).  Those described in records Y82840-Y82846 are derived
CC   from Lactadherin (BA-46).  Those described in records Y82847-Y82854
CC   are derived from Mucin and those described in records Y82871-Y82882
CC   are derived from Teratocarcinoma derived growth factor (CRIPTO-1).
XX
SQ   Sequence   9 AA;


  Query Match             54.2%;  Score 26;  DB 21;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 LSGACLNL 9
          |:| ||::
Db      2 lagiclsi 9


RESULT  14
Y49030
ID   Y49030 standard; Peptide; 7 AA.
XX
AC   Y49030;
XX
DT   10-DEC-1999  (first entry)
XX
DE   Membrane dipeptidase-binding liver homing peptide #2.
XX
KW   Homing peptide; organ; tissue; lung; pancreas; skin; retina; MDP;
KW   prostate; ovary; lymph node; adrenal gland; liver; gut; tumour;
KW   membrane dipeptidase.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9946284-A2.
XX
PD   16-SEP-1999.
XX
PF   10-MAR-1999;   99WO-US05284.
XX
```

```
PR   13-MAR-1998;   98US-0042107.
PR   26-FEB-1999;   99US-0042107.
XX
PA   (BURN-) BURNHAM INST.
XX
PI   Rajotte D,  Pasqualini R,  Ruoslahti EI;
XX
DR   WPI; 1999-571717/48.
XX
PT   New peptides which selectively home to organs or tissues, used for,
PT   e.g. identifying target ligands and for therapy of pathological
PT   conditions  -
XX
PS   Example 6; Page 156; 193pp; English.
XX
CC   The present invention describes peptides that selectively home to a
CC   tissue or organ. The peptides can be used for identifying an organ
CC   or tissue, for identifying a target molecule expressed by an organ or
CC   tissue or for treating an organ or tissue pathology, where the organ or
CC   tissue is selected from prostate, lung, skin, retina, pancreas, gut,
CC   ovary, adrenal gland, liver, and lymph node. The peptide bind to the
CC   membrane dipeptidase (MDP). Y48618 to Y49066 represent sequences
CC   which are used in the exemplification of the present invention.
XX
SQ   Sequence   7 AA;


  Query Match             50.0%;  Score 24;  DB 20;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        2 LSGACL 7
            |:| ||
Db        2 ltggcl 7


RESULT  15
Y39584
ID   Y39584 standard; peptide; 7 AA.
XX
AC   Y39584;
XX
DT   23-NOV-1999  (first entry)
XX
DE   CTLA-4 VLD CDR1 region insert.
XX
KW   CTLA-4; cytotoxic T-lymphocyte associated antigen 4; binding agent; VLD;
KW   variable-like domain; human; diagnosis; cancer; blood clot.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945110-A1.
XX
PD   10-SEP-1999.
XX
PF   05-MAR-1999;   99WO-AU00136.
XX
PR   06-MAR-1998;   98AU-0002210.
XX
PA   (DIAT-) DIATECH PTY LTD.
XX
PI   Coia G,  Galanis M,  Hudson PJ,  Irving RA,  Nuttall SD;
XX
DR   WPI; 1999-551040/46.
XX
PT   New binding agent comprising monomeric V-like domain in which at least
PT   one complementarity determining region loop is modified, useful for
PT   diagnosis of cancer  -
XX
PS   Example 6; Page 27; 117pp; English.
XX
CC   This sequence is an insert in the CDR1 of the variable-like domain (VLD)
CC   of the human cytotoxic T-lymphocyte associated antigen 4 (CTLA-4), used
CC   in the binding agent of the invention. The binding agent (I) comprises at
CC   least one monomeric VLD that is derived from a non-antibody ligand and
CC   has at least one CDR (complementarity determining region) loop sequence,
CC   or part of it, modified or replaced so that, compared to unmodified VLD,
CC   its solubility is increased and/or the size is altered and/or a
CC   disulphide bond is created within, or between, one or more CDR loops. (I)
CC   are used for diagnosis, e.g. in vivo detection/localisation of cancer,
CC   blood clots etc., also in vitro when immobilised on solid supports or
CC   biosensors and therapeutically. Modified VLD may have binding affinity
CC   for drugs, steroids, pesticides, antigens, growth factors, tumour
CC   markers, cell or viral proteins. Modification of VLD improves solubility
CC   and alters binding specificity. Since VLD are derived from human
CC   proteins, the need for a humanizing step (to avoid adverse immune
CC   responses) is avoided, and modification also improves expression in
CC   Escherichia coli.
XX
SQ   Sequence   7 AA;


  Query Match             50.0%;  Score 24;  DB 20;  Length 7;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 GACLN 8
            | |||
Db        3 grcln 7


RESULT  16
W70043
ID   W70043 standard; peptide; 9 AA.
XX
AC   W70043;
XX
DT   22-OCT-1998  (first entry)
XX
DE   MAGE 3 antigen derived HLA-A2.1 binding peptide 9 (residues 174-182).
XX
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW   antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW   viral infection; AIDS; hepatitis; bacterial infection; malaria;
KW   fungal infection; tuberculosis; melanoma; MAGE antigen.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9833888-A1.
XX
PD   06-AUG-1998.
XX
PF   30-JAN-1998;   98WO-US01959.
XX
PR   31-JAN-1997;   97US-0036696.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Celis E,  Sette A,  Sidney J,  Southwood S,  Tsai V;
XX
DR   WPI; 1998-437445/37.
XX
PT   Production of antigen-specific cytotoxic T cells - by incubating
PT   immunogenic peptide(s) from antigen that binds class I major
PT   histocompatibility complex molecules with pre-treated antigen
PT   presenting cells
XX
PS   Example 5; Page 72; 104pp; English.
XX
CC   Sequences shown in W70027 to W70043 represent peptides derived from
CC   MAGE2 and MAGE3 antigens. The peptides can bind to a human leukocyte
CC   antigen (HLA), HLA-A2.1 and are used to exemplify the method of
CC   invention of producing antigen-specific cytotoxic T cells (CTLs) in
```

```
CC   vitro. The method comprises contacting immunogenic peptides from an
CC   antigen that binds class I major histocompatibility complex (MHC)
CC   molecules with antigen presenting cells (APCs) pretreated with
CC   pretreatment growth factors, and incubating the APCs with purified CD8
CC   cells in the presence of at least 2 incubation growth factors, thereby
CC   producing antigen-specific CTLs. A method for specifically killing
CC   target cells in a human patient is also provided which comprises
CC   obtaining a fluid sample containing CTLs from a patient, contacting the
CC   cytotoxic T cells with APCs pretreated with pre-treatment growth factors,
CC   where the APCs comprise class I MHC molecules. The pretreated APCs are
CC   incubated with the cytotoxic growth factors, thereby producing activated
CC   CTLs which are contacted with a carrier to form a composition. The
CC   composition can then be administered to the patient. The activated CTLs
CC   can be used for treating cancers, immune disorders, viral infections,
CC   AIDS, hepatitis, bacterial infection, fungal infection, malaria or
CC   tuberculosis.
XX
SQ   Sequence   9 AA;

  Query Match             50.0%;  Score 24;  DB 19;  Length 9;
  Best Local Similarity   44.4%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
           |:   || |
Db       1 yifatclgl 9


RESULT  17
Y47153
ID   Y47153 standard; Peptide; 9 AA.
XX
AC   Y47153;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #1764.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 95; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
```

```
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;

  Query Match             50.0%;  Score 24;  DB 20;  Length 9;
  Best Local Similarity   44.4%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
           |:   || |
Db       1 yifatclgl 9


RESULT  18
Y47480
ID   Y47480 standard; Peptide; 9 AA.
XX
AC   Y47480;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2091.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 112; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
```

```
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;

  Query Match             50.0%;  Score 24;  DB 20;  Length 9;
  Best Local Similarity   44.4%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         1 YLSGACLNL 9
             |:   || |
Db         1 yifatclgl 9


RESULT  19
Y47562
ID   Y47562 standard; Peptide; 9 AA.
XX
AC   Y47562;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2173.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 115; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
```

```
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;

  Query Match             50.0%;  Score 24;  DB 20;  Length 9;
  Best Local Similarity   44.4%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         1 YLSGACLNL 9
             |:   || |
Db         1 yifatclgl 9


RESULT  20
Y47563
ID   Y47563 standard; Peptide; 9 AA.
XX
AC   Y47563;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2174.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 115; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
```

```
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;


  Query Match             50.0%;  Score 24;  DB 20;  Length 9;
  Best Local Similarity   44.4%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 YLSGACLNL 9
           |:   || |
Db       1 yifatclgl 9


RESULT  21
Y01944
ID   Y01944 standard; peptide; 6 AA.
XX
AC   Y01944;
XX
DT   01-JUL-1999  (first entry)
XX
DE   Fragment of the EGF-like repeat 4 of laminin.
XX
KW   Antagonist; laminin interaction; nidogen; basement membrane assembly;
KW   therapeutic agent; diabetic patient; basement membrane thickening;
KW   chronic renal failure; blindness; retinopathy; vasculitis;
KW   scleroderma; systemic lupus; tumour cell.
XX
OS   Synthetic.
XX
PN   US5493008-A.
XX
PD   20-FEB-1996.
XX
PF   15-AUG-1994;   94US-0288728.
XX
PR   15-AUG-1994;   94US-0288728.
XX
PA   (UYVI-) UNIV VIRGINIA PATENTS FOUND.
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI   Fox JW,  Timpl R;
XX
DR   WPI; 1996-150237/15.
XX
PT   Peptide antagonists for preventing laminin interaction with nidogen
PT   - useful for preventing basement membrane formation e.g. in patients
PT   with diabetes or vascular injuries
XX
PS   Disclosure; Column 11; 19pp; English.
XX
CC   The specification describes antagonists which specifically prevent
CC   laminin interaction with nidogen which is essential in the
CC   supramolecular assembly of basement membranes. These antagonists
CC   can be used as therapeutic agents in diabetic patients where basement
CC   membrane thickening leads to chronic renal failure and blindness
CC   (retinopathy). Several other vascular injuries including vasculitis,
CC   scleroderma and systemic lupus can also be treated by the antagonists.
CC   Further, the antagonists can break down basement membrane formation
CC   around tumour cells to allow attack of the cells by antibodies and
CC   immune cells. The present sequence was used in the preparation of
CC   the antagonists of the invention.
XX
SQ   Sequence   6 AA;


  Query Match             47.9%;  Score 23;  DB 17;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LSGACL 7
           |:| ||
Db       1 ltgecl 6


RESULT  22
R37279
ID   R37279 standard; Protein; 8 AA.
XX
AC   R37279;
XX
DT   06-SEP-1993  (first entry)
XX
DE   PM-1/BSA peptide.
XX
KW   Neuroendocrine; antigen; diabetes mellitis; pancrease; insulinoma;
KW   PM-1; bovine serum albumin; BSA; epitope.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1
FT                   /label= GLU, ASP
FT   Misc-difference 2
FT                   /label= GLU, LYS
FT   Misc-difference 7
FT                   /label= GLU, LEU
XX
PN   WO9309141-A.
XX
PD   13-MAY-1993.
XX
PF   29-OCT-1992;   92WO-US09428.
XX
PR   01-NOV-1991;   91US-0788118.
PR   19-JUN-1992;   92US-0901523.
XX
PA   (JOSL-) JOSLIN DIABETES CENT.
XX
PI   Eisenbarth GS,  Pietropaolo M;
XX
DR   WPI; 1993-167624/20.
XX
PT   Neuro-endocrine protein antigen PM-1 - useful for treating
PT   auto:immune diseases e.g. type I diabetes
XX
PS   Claim 20; Page 38; 56pp; English.
XX
CC   Sequence analysis of the PM-1 protein revealed two regions of
CC   similarity with bovine serum albumin (BSA).  These regions of
CC   similarity may contain epitopes shared by the PM-1 molecule and BSA.
CC   It has been shown that many patients with Type I diabetes have elevated
CC   levels of anti-IgG anti-BSA antibodies.  Thus, BSA may represent a
CC   target antigen for cow milk induced islet autoimmunity.  Peptides
CC   comprising amino acids residues shared by the PM-1 protein and BSA
CC   may be useful in the form of a therapeutic composition to treat an
CC   autoimmune disease, such as Type I diabetes in an individual.
XX
SQ   Sequence   8 AA;


  Query Match             47.9%;  Score 23;  DB 14;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       4 GACL 7
           ||||
Db       3 gacl 6


RESULT  23
R37280
ID   R37280 standard; Protein; 8 AA.
XX
AC   R37280;
XX
DT   06-SEP-1993  (first entry)
XX
DE   PM-1/BSA peptide.
XX
KW   Neuroendocrine; antigen; diabetes mellitis; pancrease; insulinoma;
KW   PM-1; bovine serum albumin; BSA; epitope.
XX
OS   Synthetic.
XX
PN   WO9309141-A.
XX
PD   13-MAY-1993.
XX
PF   29-OCT-1992;   92WO-US09428.
XX
PR   01-NOV-1991;   91US-0788118.
PR   19-JUN-1992;   92US-0901523.
XX
PA   (JOSL-) JOSLIN DIABETES CENT.
XX
PI   Eisenbarth GS,  Pietropaolo M;
XX
DR   WPI; 1993-167624/20.
XX
PT   Neuro-endocrine protein antigen PM-1 - useful for treating
PT   auto:immune diseases e.g. type I diabetes
XX
PS   Claim 20; Page 38; 56pp; English.
XX
CC   Sequence analysis of the PM-1 protein revealed two regions of
CC   similarity with bovine serum albumin (BSA).  These regions of
CC   similarity may contain epitopes shared by the PM-1 molecule and BSA.
CC   It has been shown that many patients with Type I diabetes have elevated
CC   levels of anti-IgG anti-BSA antibodies.  Thus, BSA may represent a
CC   target antigen for cow milk induced islet autoimmunity.  Peptides
CC   comprising amino acids residues shared by the PM-1 protein and BSA
CC   may be useful in the form of a therapeutic composition to treat an
CC   autoimmune disease, such as Type I diabetes in an individual.
XX
SQ   Sequence   8 AA;

  Query Match             47.9%;  Score 23;  DB 14;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GACL 7
           ||||
Db       3 gacl 6


RESULT  24
R37281
ID   R37281 standard; Protein; 8 AA.
XX
AC   R37281;
XX
DT   06-SEP-1993  (first entry)
XX
DE   PM-1/BSA peptide.
XX
```

```
KW   Neuroendocrine; antigen; diabetes mellitis; pancrease; insulinoma;
KW   PM-1; bovine serum albumin; BSA; epitope.
XX
OS   Synthetic.
XX
PN   WO9309141-A.
XX
PD   13-MAY-1993.
XX
PF   29-OCT-1992;   92WO-US09428.
XX
PR   01-NOV-1991;   91US-0788118.
PR   19-JUN-1992;   92US-0901523.
XX
PA   (JOSL-) JOSLIN DIABETES CENT.
XX
PI   Eisenbarth GS,  Pietropaolo M;
XX
DR   WPI; 1993-167624/20.
XX
PT   Neuro-endocrine protein antigen PM-1 - useful for treating
PT   auto:immune diseases e.g. type I diabetes
XX
PS   Claim 20; Page 38; 56pp; English.
XX
CC   Sequence analysis of the PM-1 protein revealed two regions of
CC   similarity with bovine serum albumin (BSA).  These regions of
CC   similarity may contain epitopes shared by the PM-1 molecule and BSA.
CC   It has been shown that many patients with Type I diabetes have elevated
CC   levels of anti-IgG anti-BSA antibodies.  Thus, BSA may represent a
CC   target antigen for cow milk induced islet autoimmunity.  Peptides
CC   comprising amino acids residues shared by the PM-1 protein and BSA
CC   may be useful in the form of a therapeutic composition to treat an
CC   autoimmune disease, such as Type I diabetes in an individual.
XX
SQ   Sequence   8 AA;

  Query Match             47.9%;  Score 23;  DB 14;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GACL 7
           ||||
Db       3 gacl 6


RESULT  25
R37282
ID   R37282 standard; Protein; 8 AA.
XX
AC   R37282;
XX
DT   06-SEP-1993  (first entry)
XX
DE   PM-1/BSA peptide.
XX
KW   Neuroendocrine; antigen; diabetes mellitis; pancrease; insulinoma;
KW   PM-1; bovine serum albumin; BSA; epitope.
XX
OS   Synthetic.
XX
PN   WO9309141-A.
XX
PD   13-MAY-1993.
XX
PF   29-OCT-1992;   92WO-US09428.
XX
PR   01-NOV-1991;   91US-0788118.
PR   19-JUN-1992;   92US-0901523.
XX
PA   (JOSL-) JOSLIN DIABETES CENT.
XX
```

```
PI   Eisenbarth GS,  Pietropaolo M;
XX
DR   WPI; 1993-167624/20.
XX
PT   Neuro-endocrine protein antigen PM-1 - useful for treating
PT   auto:immune diseases e.g. type I diabetes
XX
PS   Claim 20; Page 38; 56pp; English.
XX
CC   Sequence analysis of the PM-1 protein revealed two regions of
CC   similarity with bovine serum albumin (BSA).  These regions of
CC   similarity may contain epitopes shared by the PM-1 molecule and BSA.
CC   It has been shown that many patients with Type I diabetes have elevated
CC   levels of anti-IgG anti-BSA antibodies.  Thus, BSA may represent a
CC   target antigen for cow milk induced islet autoimmunity.  Peptides
CC   comprising amino acids residues shared by the PM-1 protein and BSA
CC   may be useful in the form of a therapeutic composition to treat an
CC   autoimmune disease, such as Type I diabetes in an individual.
XX
SQ   Sequence   8 AA;

  Query Match             47.9%;  Score 23;  DB 14;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GACL 7
           ||||
Db       3 gacl 6


RESULT  26
R37283
ID   R37283 standard; Protein; 8 AA.
XX
AC   R37283;
XX
DT   06-SEP-1993  (first entry)
XX
DE   PM-1/BSA peptide.
XX
KW   Neuroendocrine; antigen; diabetes mellitis; pancrease; insulinoma;
KW   PM-1; bovine serum albumin; BSA; epitope.
XX
OS   Synthetic.
XX
PN   WO9309141-A.
XX
PD   13-MAY-1993.
XX
PF   29-OCT-1992;   92WO-US09428.
XX
PR   01-NOV-1991;   91US-0788118.
PR   19-JUN-1992;   92US-0901523.
XX
PA   (JOSL-) JOSLIN DIABETES CENT.
XX
PI   Eisenbarth GS,  Pietropaolo M;
XX
DR   WPI; 1993-167624/20.
XX
PT   Neuro-endocrine protein antigen PM-1 - useful for treating
PT   auto:immune diseases e.g. type I diabetes
XX
PS   Claim 20; Page 38; 56pp; English.
XX
CC   Sequence analysis of the PM-1 protein revealed two regions of
CC   similarity with bovine serum albumin (BSA).  These regions of
CC   similarity may contain epitopes shared by the PM-1 molecule and BSA.
CC   It has been shown that many patients with Type I diabetes have elevated
CC   levels of anti-IgG anti-BSA antibodies.  Thus, BSA may represent a
CC   target antigen for cow milk induced islet autoimmunity.  Peptides
CC   comprising amino acids residues shared by the PM-1 protein and BSA
CC   may be useful in the form of a therapeutic composition to treat an
CC   autoimmune disease, such as Type I diabetes in an individual.
XX
SQ   Sequence   8 AA;

  Query Match             47.9%;  Score 23;  DB 14;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GACL 7
           ||||
Db       3 gacl 6


RESULT  27
R37284
ID   R37284 standard; Protein; 8 AA.
XX
AC   R37284;
XX
DT   06-SEP-1993  (first entry)
XX
DE   PM-1/BSA peptide.
XX
KW   Neuroendocrine; antigen; diabetes mellitis; pancrease; insulinoma;
KW   PM-1; bovine serum albumin; BSA; epitope.
XX
OS   Synthetic.
XX
PN   WO9309141-A.
XX
PD   13-MAY-1993.
XX
PF   29-OCT-1992;   92WO-US09428.
XX
PR   01-NOV-1991;   91US-0788118.
PR   19-JUN-1992;   92US-0901523.
XX
PA   (JOSL-) JOSLIN DIABETES CENT.
XX
PI   Eisenbarth GS,  Pietropaolo M;
XX
DR   WPI; 1993-167624/20.
XX
PT   Neuro-endocrine protein antigen PM-1 - useful for treating
PT   auto:immune diseases e.g. type I diabetes
XX
PS   Claim 20; Page 38; 56pp; English.
XX
CC   Sequence analysis of the PM-1 protein revealed two regions of
CC   similarity with bovine serum albumin (BSA).  These regions of
CC   similarity may contain epitopes shared by the PM-1 molecule and BSA.
CC   It has been shown that many patients with Type I diabetes have elevated
CC   levels of anti-IgG anti-BSA antibodies.  Thus, BSA may represent a
CC   target antigen for cow milk induced islet autoimmunity.  Peptides
CC   comprising amino acids residues shared by the PM-1 protein and BSA
CC   may be useful in the form of a therapeutic composition to treat an
CC   autoimmune disease, such as Type I diabetes in an individual.
XX
SQ   Sequence   8 AA;

  Query Match             47.9%;  Score 23;  DB 14;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GACL 7
           ||||
Db       3 gacl 6


RESULT  28
```

```
R37285
ID   R37285 standard; Protein; 8 AA.
XX
AC   R37285;
XX
DT   06-SEP-1993  (first entry)
XX
DE   PM-1/BSA peptide.
XX
KW   Neuroendocrine; antigen; diabetes mellitis; pancrease; insulinoma;
KW   PM-1; bovine serum albumin; BSA; epitope.
XX
OS   Synthetic.
XX
PN   WO9309141-A.
XX
PD   13-MAY-1993.
XX
PF   29-OCT-1992;   92WO-US09428.
XX
PR   01-NOV-1991;   91US-0788118.
PR   19-JUN-1992;   92US-0901523.
XX
PA   (JOSL-) JOSLIN DIABETES CENT.
XX
PI   Eisenbarth GS,  Pietropaolo M;
XX
DR   WPI; 1993-167624/20.
XX
PT   Neuro-endocrine protein antigen PM-1 - useful for treating
PT   auto:immune diseases e.g. type I diabetes
XX
PS   Claim 20; Page 38; 56pp; English.
XX
CC   Sequence analysis of the PM-1 protein revealed two regions of
CC   similarity with bovine serum albumin (BSA).  These regions of
CC   similarity may contain epitopes shared by the PM-1 molecule and BSA.
CC   It has been shown that many patients with Type I diabetes have elevated
CC   levels of anti-IgG anti-BSA antibodies.  Thus, BSA may represent a
CC   target antigen for cow milk induced islet autoimmunity.  Peptides
CC   comprising amino acids residues shared by the PM-1 protein and BSA
CC   may be useful in the form of a therapeutic composition to treat an
CC   autoimmune disease, such as Type I diabetes in an individual.
XX
SQ   Sequence   8 AA;


  Query Match             47.9%;  Score 23;  DB 14;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 GACL 7
          ||||
Db      3 gacl 6


RESULT  29
R37286
ID   R37286 standard; Protein; 8 AA.
XX
AC   R37286;
XX
DT   06-SEP-1993  (first entry)
XX
DE   PM-1/BSA peptide.
XX
KW   Neuroendocrine; antigen; diabetes mellitis; pancrease; insulinoma;
KW   PM-1; bovine serum albumin; BSA; epitope.
XX
OS   Synthetic.
XX
PN   WO9309141-A.
XX
```

```
PD   13-MAY-1993.
XX
PF   29-OCT-1992;   92WO-US09428.
XX
PR   01-NOV-1991;   91US-0788118.
PR   19-JUN-1992;   92US-0901523.
XX
PA   (JOSL-) JOSLIN DIABETES CENT.
XX
PI   Eisenbarth GS,  Pietropaolo M;
XX
DR   WPI; 1993-167624/20.
XX
PT   Neuro-endocrine protein antigen PM-1 - useful for treating
PT   auto:immune diseases e.g. type I diabetes
XX
PS   Claim 20; Page 38; 56pp; English.
XX
CC   Sequence analysis of the PM-1 protein revealed two regions of
CC   similarity with bovine serum albumin (BSA).  These regions of
CC   similarity may contain epitopes shared by the PM-1 molecule and BSA.
CC   It has been shown that many patients with Type I diabetes have elevated
CC   levels of anti-IgG anti-BSA antibodies.  Thus, BSA may represent a
CC   target antigen for cow milk induced islet autoimmunity.  Peptides
CC   comprising amino acids residues shared by the PM-1 protein and BSA
CC   may be useful in the form of a therapeutic composition to treat an
CC   autoimmune disease, such as Type I diabetes in an individual.
XX
SQ   Sequence   8 AA;


  Query Match             47.9%;  Score 23;  DB 14;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 GACL 7
          ||||
Db      3 gacl 6


RESULT  30
R37287
ID   R37287 standard; Protein; 8 AA.
XX
AC   R37287;
XX
DT   06-SEP-1993  (first entry)
XX
DE   PM-1/BSA peptide.
XX
KW   Neuroendocrine; antigen; diabetes mellitis; pancrease; insulinoma;
KW   PM-1; bovine serum albumin; BSA; epitope.
XX
OS   Synthetic.
XX
PN   WO9309141-A.
XX
PD   13-MAY-1993.
XX
PF   29-OCT-1992;   92WO-US09428.
XX
PR   01-NOV-1991;   91US-0788118.
PR   19-JUN-1992;   92US-0901523.
XX
PA   (JOSL-) JOSLIN DIABETES CENT.
XX
PI   Eisenbarth GS,  Pietropaolo M;
XX
DR   WPI; 1993-167624/20.
XX
PT   Neuro-endocrine protein antigen PM-1 - useful for treating
PT   auto:immune diseases e.g. type I diabetes
XX
```

```
PS   Claim 20; Page 38; 56pp; English.
XX
CC   Sequence analysis of the PM-1 protein revealed two regions of
CC   similarity with bovine serum albumin (BSA).  These regions of
CC   similarity may contain epitopes shared by the PM-1 molecule and BSA.
CC   It has been shown that many patients with Type I diabetes have elevated
CC   levels of anti-IgG anti-BSA antibodies.  Thus, BSA may represent a
CC   target antigen for cow milk induced islet autoimmunity.  Peptides
CC   comprising amino acids residues shared by the PM-1 protein and BSA
CC   may be useful in the form of a therapeutic composition to treat an
CC   autoimmune disease, such as Type I diabetes in an individual.
XX
SQ   Sequence   8 AA;


  Query Match             47.9%;  Score 23;  DB 14;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        4 GACL 7
            ||||
Db        3 gacl 6


RESULT  31
W13449
ID   W13449 standard; Peptide; 8 AA.
XX
AC   W13449;
XX
DT   15-JAN-1998  (first entry)
XX
DE   Kidney homing peptide.
XX
KW   Kidney homing peptide; in vivo panning; screening; phage display;
KW   drug delivery.
XX
OS   Synthetic.
XX
PN   WO9710507-A1.
XX
PD   20-MAR-1997.
XX
PF   10-SEP-1996;   96WO-US14600.
XX
PR   11-SEP-1995;   95US-0526710.
PR   11-SEP-1995;   95US-0526708.
XX
PA   (LJOL-) LA JOLLA CANCER RES FOUND.
XX
PI   Pasqualini R,  Ruoslahti E;
XX
DR   WPI; 1997-202359/18.
XX
PT   Obtaining compound that homes to selected organ or tissue - by in
PT   vivo panning method, specifically to identify brain, kidney,
PT   angiogenic vasculature or tumour tissue homing peptide(s)
XX
PS   Example 2; Page 55; 75pp; English.
XX
CC   This synthetic peptide is an example of a kidney-homing peptide
CC   that was identified using a claimed method for obtaining
CC   molecules that home to a selected organ or tissue.  This in vivo
CC   panning method typically involves administering a phage display
CC   library to a subject, and identifying expressed peptides which
CC   home to the desired organ or tissue, e.g. brain, kidney, angiogenic
CC   vascular tissue or tumour tissue.  The isolated peptides (see
CC   W13412-52, W11181-86) can be used to target e.g. drugs, toxins or
CC   labels to the selected organ/tissue (claimed) or to identify and/or
CC   isolate target molecules (claimed).  The peptides can be directly
CC   identified in vivo, as compared to prior art in vitro screening
CC   methods, which require further examination to see if they maintain
CC   specificity in vivo.
```

```
XX
SQ   Sequence   8 AA;


  Query Match             47.9%;  Score 23;  DB 18;  Length 8;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        2 LSGAC 6
            ||| |
Db        4 lsgvc 8


RESULT  32
W77008
ID   W77008 standard; peptide; 8 AA.
XX
AC   W77008;
XX
DT   09-NOV-1998  (first entry)
XX
DE   Peptide mimetic of cytokine receptor gamma chain 31.
XX
KW   Peptide mimetic; gamma-chain; cytokine receptor; signal transduction;
KW   autoimmune disease; graft vs. host disease; transplant rejection;
KW   graft rejection; interleukin; immunosuppressant; T cell; B cell.
XX
OS   Synthetic.
XX
PN   WO9834631-A1.
XX
PD   13-AUG-1998.
XX
PF   06-FEB-1998;   98WO-US02339.
XX
PR   07-FEB-1997;   97US-0036941.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Korngold R,  Townsend RM;
XX
DR   WPI; 1998-446944/38.
XX
PT   New peptide(s) mimicking a loop in the gamma chain of cytokine
PT   receptors - inhibit signal transduction through these receptors,
PT   useful as immunosuppressants for treating or preventing e.g.
PT   leukaemia, autoimmune disease, graft rejection
XX
PS   Claim 8; Page 29; 54pp; English.
XX
CC   The peptide mimetics W61734-W61756 and W77001-W77014 mimic a loop on the
CC   gamma-chain of cytokine receptors, and interacts with a cytokine or a
CC   gamma-chain partner receptor chain of a heterodimeric cytokine receptor.
CC   They inhibit signal transduction mediated by cytokine:receptor binding
CC   (of cytokines that bind to receptors with a gamma-chain).  They are used
CC   to inhibit or suppress cytokine-mediated immune responses, growth,
CC   proliferation, function and activity of cells. Particularly they are used
CC   for treatment or prevention of lymphoma, leukaemia, allergy
CC   (e.g. asthma), autoimmune disease (e.g. rheumatoid arthritis, lupus,
CC   multiple sclerosis or myasthenia gravis), graft vs. host disease and
CC   transplant or graft rejection.  They inhibit function of interleukin
CC   (IL)-2, -4, -7, -9, -13 and -15 (the receptors for which include a common
CC   gamma -chain), so function as immunosuppressants by reducing
CC   proliferation of T and B cells.
XX
SQ   Sequence   8 AA;


  Query Match             47.9%;  Score 23;  DB 19;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 CLNL 9
```

```
               ||||
Db          1 clnl 4


RESULT  33
W40267
ID   W40267 standard; Protein; 9 AA.
XX
AC   W40267;
XX
DT   16-JUN-1998  (first entry)
XX
DE   K. oxytoca R-specific amidohydrolase peptide T5.
XX
KW   R-specific amidohydrolase; hydrolysis; racemate; nitrogen source;
KW   3,3,3-trifluoro-2-hydroxy-2-methyl propionamide; isoform.
XX
OS   Klebsiella oxytoca.
XX
PN   WO9801568-A2.
XX
PD   15-JAN-1998.
XX
PF   10-JUL-1997;   97WO-EP03670.
XX
PR   03-MAR-1997;   97CH-0000500.
PR   10-JUL-1996;   96CH-0001723.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Brieden W,  Naughton A,  Robins K,  Shaw N,  Tinschert A;
PI   Zimmermann T;
XX
DR   WPI; 1998-101063/09.
XX
PT   Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation
PT   - by stereoselective hydrolysis of corresponding racemic amide using
PT   microorganism or derived enzyme, used as drug intermediate
XX
PS   Example 10.2; Page 29; 68pp; German.
XX
CC   W40264-W40275 are peptide fragments of an R-specific amidohydrolase
CC   isolated from Klebsiella oxytoca strain PRS1. This enzyme allows the
CC   microorganism to utilise 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
CC   as its sole nitrogen source. This amidohydrolase is used in a process for
CC   preparing (R)-isoforms of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
CC   which is cheaper than prior art optical resolution of the racemate using
CC   dimethoxy strychnine or (S)-(-)- alpha -methylbenzylamine.
XX
SQ   Sequence   9 AA;


  Query Match             47.9%;  Score 23;  DB 19;  Length 9;
  Best Local Similarity   62.5%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 YLSGACLN 8
              |  || ||
Db          1 ytvgamln 8


RESULT  34
Y82878
ID   Y82878 standard; peptide; 9 AA.
XX
AC   Y82878;
XX
DT   19-JUN-2000  (first entry)
XX
DE   Teratocarcinoma-derived growth factor (CRIPTO-1) antigenic peptide.
XX
KW   Tumour associated antigen peptide; TAA; cancer; carcinoma;
KW   treatment; prevention; cure; anti-tumour vaccine; metastases;
KW   breast; bladder; prostate; pancreas; ovary; thyroid; colon;
KW   stomach; carcinoma; MHC Class I; HLA-A2; human;
KW   Major Histocompatibility Complex; uroplakin;
KW   prostate specific antigen; prostate specific membrane antigen;
KW   prostate acid phosphatase; mucin; lactadherin;
KW   teratocarcinoma derived growth factor; PSA; PSMA; PAP; CRIPTO-1.
XX
OS   Homo sapiens.
XX
PN   WO200006723-A1.
XX
PD   10-FEB-2000.
XX
PF   29-JUL-1999;   99WO-IL00417.
XX
PR   30-JUL-1998;   98IL-0125608.
XX
PA   (YEDA ) YEDA RES & DEV CO LTD.
PA   (BIOT-) BIO-TECHNOLOGY GEN CORP.
XX
PI   Eisenbach L,  Carmon L,  Tirosh B,  Bar-haim E,  Paz A,  Fridkin M;
PI   Fitzer-attas C;
XX
DR   WPI; 2000-205463/18.
XX
PT   Tumor associated antigen peptides, especially derived from uroplakin,
PT   useful as vaccines to prevent or cure cancers including breast,
PT   bladder, prostate, pancreas, ovary, thyroid, colon and stomach -
XX
PS   Claim 19; Page 108; 113pp; English.
XX
CC   Tumour associated antigen peptides (TAA) may be used for the
CC   treatment, prevention and cure of cancer or cancer metastases. The
CC   cancer may be breast, bladder, prostate, pancreas, ovary, thyroid,
CC   colon, stomach, head or neck cancer or a carcinoma. The tumour
CC   associated antigens are presentable to the immune system by HLA-A2
CC   molecules and are generally between 8 to 10 amino acids in length.
CC   The amino acids located at positions 2 and 9 of the tumour associated
CC   antigens are the anchor residues which participate in the binding to
CC   MHC class I molecules, more specifically HLA-A2.  More tumour
CC   associated antigens are described in GENESEQ records Y82806-Y82882.
CC   Those tumour associated antigens described in records Y82806-Y82824
CC   and Y82855-YY82869 are derived from Uroplakin, such as Uroplakin II,
CC   Uroplakin Ia, Uroplakin III and Uroplakin Ib.  Those described in
CC   records Y82825-Y82829 are derived from prostate specific antigen
CC   (PSA). Those described in records Y82830-Y82835 are derived from
CC   prostate specific membrane antigen (PSMA).  Those described in
CC   records YY82836-Y82839 are derived from prostate acid phosphotase
CC   (PAP).  Those described in records Y82840-Y82846 are derived
CC   from Lactadherin (BA-46).  Those described in records Y82847-Y82854
CC   are derived from Mucin and those described in records Y82871-Y82882
CC   are derived from Teratocarcinoma derived growth factor (CRIPTO-1).
XX
SQ   Sequence   9 AA;


  Query Match             47.9%;  Score 23;  DB 21;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 LSGACLNL 9
              | | ||::
Db          2 lvgiclsi 9


RESULT  35
W49470
ID   W49470 standard; peptide; 9 AA.
XX
AC   W49470;
XX
DT   05-JUN-1998  (first entry)
XX
```

```
DE   Human leucocyte antigen DQ4 binding peptide #361.
XX
KW   Human leucocyte antigen; HLA-DQ4; combinatorial library;
KW   autoimmune disease; chronic articular rheumatism.
XX
OS   Synthetic.
XX
PN   JP08151396-A.
XX
PD   11-JUN-1996.
XX
PF   28-NOV-1994;   94JP-0292657.
XX
PR   28-NOV-1994;   94JP-0292657.
XX
PA   (TEIJ ) TEIJIN LTD.
XX
DR   WPI; 1996-329479/33.
XX
PT   HLA-binding oligopeptide and an immuno:regulator contg it - used in
PT   the treatment of auto:immune disease
XX
PS   Claim 4; Page 42; 61pp; Japanese.
XX
CC   This peptide is an example of a peptide which binds to a human leucocyte
CC   antigen HLA-DQ4 molecule.  The peptide was isolated from a phagemid
CC   combinatorial library comprising the sequence V05953, by screening with
CC   an HLA-DQ4 molecule.  The peptide is used for the treatment of autoimmune
CC   disease, or especially for treatment of viral diseases.
XX
SQ   Sequence   9 AA;

  Query Match             45.8%;  Score 22;  DB 17;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 SGACLNL 9
           :|||  |
Db       1 agacrgl 7


Search completed: December 16, 2000, 03:07:19
Job time: 8060 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 02:00:57 ; Search time 113.2 Seconds
                                              (without alignments)
                                              7.424 Million cell updates/sec

Title:           US-09-529-121-4
Perfect score:   45
Sequence:        1 YLSGANINL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters:      297973

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_14:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                  %
Result          Query
  No.   Score   Match Length DB  ID                    Description
--------------------------------------------------------------------------
    1      36    80.0    131  8  Q35903                Q35903 strongyloce
    2      36    80.0    554  2  O07411                O07411 mycobacteri
    3      35    77.8    143  4  Q9UPC0                Q9upc0 homo sapien
    4      35    77.8    180  4  Q14628                Q14628 homo sapien
    5      34    75.6    385  2  P94599                P94599 bacillus th
    6      34    75.6    790  4  Q9Y2D7                Q9y2d7 homo sapien
    7      34    75.6    877  5  Q9VT60                Q9vt60 drosophila
    8      34    75.6    949  2  Q9RTF5                Q9rtf5 deinococcus
    9      34    75.6    979  5  Q21962                Q21962 caenorhabdi
   10      34    75.6    985  5  Q9VH09                Q9vh09 drosophila
   11      34    75.6   1024 13  Q9PUU9                 Q9puu9 anas platyr
   12      33    73.3    298 12  Q9WA61                 Q9wa61 turnip mosa
   13      33    73.3    507  2  O84150                O84150 chlamydia t
   14      33    73.3    635  2  P94146                P94146 alcaligenes
   15      33    73.3   1321  1  O26339                O26339 methanobact
   16      33    73.3   1956  5  Q20641                Q20641 caenorhabdi
   17      32    71.1    236  5  Q20421                Q20421 caenorhabdi
   18      32    71.1    248  5  Q9XWR6                Q9xwr6 caenorhabdi
   19      32    71.1    259  2  Q55773                Q55773 synechocyst
   20      32    71.1    272  5  Q9XVM8                Q9xvm8 caenorhabdi
   21      32    71.1    311  5  O01675                O01675 branchiosto
   22      32    71.1    343  5  O01758                O01758 caenorhabdi
   23      32    71.1    349 12  Q65125                 Q65125 african swi
   24      32    71.1    356  2  Q9WYJ4                Q9wyj4 thermotoga
   25      32    71.1    447 10  Q9SLV4                 Q9slv4 pisum sativ
   26      32    71.1    473  2  Q9RB20                Q9rb20 pectobacter
   27      32    71.1    555  4  Q9Y625                Q9y625 homo sapien
   28      32    71.1    555 11  Q9R087                 Q9r087 mus musculu
   29      32    71.1    693 12  Q82041                 Q82041 human rotav
   30      32    71.1    702  5  Q9VF80                Q9vf80 drosophila
   31      32    71.1    753  5  O02312                O02312 caenorhabdi
   32      32    71.1    952  2  O32915                O32915 mycobacteri
   33      32    71.1    983  2  P74416                P74416 synechocyst
   34      32    71.1   1031 10  O22575                 O22575 hordeum spo
   35      32    71.1   1032 10  O38766                 O38766 avena sativ
   36      32    71.1   1037 10  O82642                 O82642 arabidopsis
   37      32    71.1   1044 10  O80988                 O80988 arabidopsis
   38      32    71.1   1876  2  Q9R686                Q9r686 bacillus su
   39      32    71.1   4450  2  Q44928                Q44928 bacillus br
   40      31    68.9    104  9  O80257                O80257 vibrio chol
   41      31    68.9    266  2  Q9ZJX8                Q9zjx8 helicobacte
   42      31    68.9    285  2  P73419                P73419 synechocyst
   43      31    68.9    336 10  Q9ST82                 Q9st82 oryza sativ
   44      31    68.9    355 12  Q65122                 Q65122 african swi
   45      31    68.9    356  5  O44592                O44592 caenorhabdi
```

```
                                ALIGNMENTS


RESULT   1
Q35903
ID   Q35903      PRELIMINARY;      PRT;   131 AA.
AC   Q35903;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   NADH DEHYDROGENASE SUBUNIT 5 (ND5) (FRAGMENT).
OS   Strongylocentrotus pallidus.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
OC   Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91175684.
RA   Thomas W.K., Maa J., Wilson A.C.;
RT   "Shifting constraints on tRNA genes during mitochondrial DNA evolution
RT   in animals.";
RL   New Biol. 1:93-100(1989).
DR   EMBL; M27524; CAB25461.1; -.
KW   Mitochondrion.
FT   NON_TER     131     131
SQ   SEQUENCE   131 AA;  14405 MW;  903318CDAD6E4C0D CRC64;


  Query Match             80.0%;  Score 36;  DB 8;  Length 131;
  Best Local Similarity   77.8%;  Pred. No. 6.1;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
             ||| :||||
Db      31 YLSNSNINL 39


RESULT   2
O07411
ID   O07411      PRELIMINARY;      PRT;   554 AA.
AC   O07411;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
```

```
DE   PROBABLE FATTY-ACID COA LIGASE.
GN   RV0166 OR MTCI28.06.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 98295987.
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Sqares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
DR   EMBL; Z97050; CAB09749.1; -.
DR   HSSP; P08659; 1LCI.
DR   TUBERCULIST; Rv0166; -.
DR   INTERPRO; IPR000873; -.
DR   PFAM; PF00501; AMP-binding; 1.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Ligase.
SQ   SEQUENCE   554 AA;  59905 MW;  3AC3100FAF0B9E88 CRC64;


  Query Match             80.0%;  Score 36;  DB 2;  Length 554;
  Best Local Similarity   87.5%;  Pred. No. 31;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANIN 8
           | ||||||
Db     217 YTSGANIN 224


RESULT   3
Q9UPC0
ID   Q9UPC0      PRELIMINARY;      PRT;   143 AA.
AC   Q9UPC0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   INTERLEUKIN-1 INTRACELLULAR RECEPTOR ANTAGONIST VARIANT (FRAGMENT).
GN   IL1RN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98183404.
RA   Weissbach L., Tran K., Colquhoun S.A., Champliaud M.F., Towle C.A.;
RT   "Detection of an interleukin-1 intracellular receptor antagonist mRNA
RT   variant.";
RL   Biochem. Biophys. Res. Commun. 244:91-95(1998).
DR   EMBL; AF043143; AAC39672.1; -.
DR   INTERPRO; IPR000975; -.
DR   PFAM; PF00340; interleukin-1; 1.
DR   PRINTS; PR00264; INTERLEUKIN1.
DR   PROSITE; PS00253; INTERLEUKIN_1; 1.
KW   Receptor.
FT   NON_TER     143    143
SQ   SEQUENCE   143 AA;  16142 MW;  4CAD6784B890906B CRC64;


  Query Match             77.8%;  Score 35;  DB 4;  Length 143;
  Best Local Similarity   66.7%;  Pred. No. 11;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy       1 YLSGANINL 9
           || | |:||
Db      25 YLQGPNVNL 33


RESULT   4
Q14628
ID   Q14628      PRELIMINARY;      PRT;   180 AA.
AC   Q14628;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   INTRACELLULAR IL-1 RECEPTOR ANTAGONIST TYPE II.
GN   IL-1RN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95355865.
RA   Muzio M., Polentarutti N., Sironi M., Poli G., De Gioia L.,
RA   Introna M., Mantovani A., Colotta F.;
RT   "Cloning and characterization of a new isoform of the interleukin 1
RT   receptor antagonist.";
RL   J. Exp. Med. 182:623-628(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97146044.
RA   Jenkins J.K., Drong R.F., Shuck M.E., Bienkowski M.J., Slightom J.L.,
RA   Arend W.P., Smith M.F. Jr.;
RT   "Intracellular IL-1 receptor antagonist promoter: cell type-specific
RT   and inducible regulatory regions.";
RL   J. Immunol. 158:748-755(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Slightom J.L.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X84348; CAA59087.1; -.
DR   EMBL; U65590; AAB92269.1; -.
DR   HSSP; P18510; 1IRP.
DR   INTERPRO; IPR000975; -.
DR   PFAM; PF00340; interleukin-1; 1.
DR   PRINTS; PR00264; INTERLEUKIN1.
DR   PROSITE; PS00253; INTERLEUKIN_1; 1.
SQ   SEQUENCE   180 AA;  19897 MW;  624A1574C2334229 CRC64;


  Query Match             77.8%;  Score 35;  DB 4;  Length 180;
  Best Local Similarity   66.7%;  Pred. No. 14;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
           || | |:||
Db      62 YLQGPNVNL 70


RESULT   5
P94599
ID   P94599      PRELIMINARY;      PRT;   385 AA.
AC   P94599;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAY-1997 (TrEMBLrel. 03, Last annotation update)
DE   SERINE/THREONINE KINASE.
GN   PK1.
OS   Bacillus thuringiensis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=84-I-1-13;
RA   Dunn M.G., Ellar D.J.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
```

```
DR   EMBL; Y10168; CAA71249.1; -.
SQ   SEQUENCE   385 AA;  43949 MW;  6013163292D0329E CRC64;


  Query Match             75.6%;  Score 34;  DB 2;  Length 385;
  Best Local Similarity   75.0%;  Pred. No. 52;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGANIN 8
              |||| |:|
Db        278 YLSGPNLN 285


RESULT   6
Q9Y2D7
ID   Q9Y2D7       PRELIMINARY;      PRT;   790 AA.
AC   Q9Y2D7;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   KIAA0922 PROTEIN (FRAGMENT).
GN   KIAA0922.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 99246063.
RA   Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirosawa M.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIII.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:63-70(1999).
DR   EMBL; AB023139; BAA76766.1; -.
DR   INTERPRO; IPR002465; -.
DR   PROSITE; PS00340; RECEPTOR_CYTOKINES_2; UNKNOWN_1.
FT   NON_TER       1      1
SQ   SEQUENCE   790 AA;  87995 MW;  24F70D670D70F946 CRC64;


  Query Match             75.6%;  Score 34;  DB 4;  Length 790;
  Best Local Similarity   66.7%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              | || |:||
Db        215 YASGINVNL 223


RESULT   7
Q9VT60
ID   Q9VT60       PRELIMINARY;      PRT;   877 AA.
AC   Q9VT60;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CG6718 PROTEIN.
GN   CG6718.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE; 20196006.
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003550; AAF50194.1; -.
DR   FLYBASE; FBgn0036053; CG6718.
DR   INTERPRO; IPR002110; -.
DR   PFAM; PF00023; ank; 5.
SQ   SEQUENCE   877 AA;  96862 MW;  C9DC2CD6C282869B CRC64;


  Query Match             75.6%;  Score 34;  DB 5;  Length 877;
  Best Local Similarity   75.0%;  Pred. No. 1.3e+02;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LSGANINL 9
              |:|||:||
Db        240 LAGANVNL 247


RESULT   8
Q9RTF5
ID   Q9RTF5       PRELIMINARY;      PRT;   949 AA.
AC   Q9RTF5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   GLYCINE CLEAVAGE SYSTEM P PROTEIN.
GN   DR1809.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE; 20036896.
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome Sequence of the Radioresistant Bacterium Deinococcus
RT   radiodurans R1.";
```

```
RL   Science 286:1571-1577(1999).
DR   EMBL; AE002021; AAF11360.1; -.
DR   TIGR; DR1809; -.
SQ   SEQUENCE   949 AA;  102122 MW;  FDCBA42D4E0888D5 CRC64;

  Query Match             75.6%;  Score 34;  DB 2;  Length 949;
  Best Local Similarity   75.0%;  Pred. No. 1.4e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANIN 8
          || |||:|
Db    668 YLDGANMN 675


RESULT   9
Q21962
ID   Q21962      PRELIMINARY;      PRT;   979 AA.
AC   Q21962;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   SIMILAR TO GLYCINE DEHYDROGENASE.
GN   R12C12.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Favello T.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U23510; AAC46780.1; -.
SQ   SEQUENCE   979 AA;  108865 MW;  F445CA0B46C7AA02 CRC64;

  Query Match             75.6%;  Score 34;  DB 5;  Length 979;
  Best Local Similarity   75.0%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANIN 8
          || |||:|
Db    702 YLDGANMN 709


RESULT  10
Q9VH09
ID   Q9VH09      PRELIMINARY;      PRT;   985 AA.
AC   Q9VH09;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG3999 PROTEIN.
GN   CG3999.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE; 20196006.
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003686; AAF54512.1; -.
DR   FLYBASE; FBgn0037801; CG3999.
SQ   SEQUENCE   985 AA;  109713 MW;  F8FBF5CFD251EDE7 CRC64;

  Query Match             75.6%;  Score 34;  DB 5;  Length 985;
  Best Local Similarity   75.0%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANIN 8
          || |||:|
Db    701 YLDGANMN 708


RESULT  11
Q9PUU9
ID   Q9PUU9      PRELIMINARY;      PRT;  1024 AA.
AC   Q9PUU9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   GLYCINE DECARBOXYLASE P PROTEIN.
OS   Anas platyrhynchos (Domestic duck).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEKIN DUCK; TISSUE=LIVER;
RX   MEDLINE; 96323119.
RA   Li J.S., Tong S.P., Wands J.R.;
RT   "Characterization of a 120-Kilodalton pre-S-binding protein as a
RT   candidate duck hepatitis B virus receptor.";
RL   J. Virol. 70:6029-6035(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEKIN DUCK; TISSUE=LIVER;
RA   Li J., Tong S., Wands J.R.;
RT   "Identification and Expression of Glycine Decarboxylase (p120) as a
RT   Duck Hepatitis B Virus Pre-S Envelope-binding Protein.";
RL   J. Biol. Chem. 274:27658-27665(1999).
DR   EMBL; AF137264; AAD56281.1; -.
SQ   SEQUENCE   1024 AA;  113473 MW;  657EFC89374DFEC3 CRC64;


  Query Match             75.6%;  Score 34;  DB 13;  Length 1024;
  Best Local Similarity   75.0%;  Pred. No. 1.6e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANIN 8
             || |||:|
Db     729 YLDGANMN 736


RESULT  12
Q9WA61
ID   Q9WA61      PRELIMINARY;      PRT;   298 AA.
AC   Q9WA61;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   POLYPROTEIN (FRAGMENT).
OS   Turnip mosaic virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
OC   Potyvirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R9-10;
RA   Sohn S.-H., Choi H.-S., Jeon W.-B., Ryu T.-H., Hwang Y.-S.;
RT   "The coat protein gene of Turnip Mosaic Virus isolated from Radish in
RT   Korea.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF103785; AAD20296.1; -.
DR   INTERPRO; IPR001592; -.
DR   PFAM; PF00767; Poty_coat; 1.
KW   Polyprotein.
FT   NON_TER       1      1
FT   CHAIN        11    298       COAT PROTEIN.
SQ   SEQUENCE   298 AA;  34124 MW;  44AAD1DEC41498F8 CRC64;


  Query Match             73.3%;  Score 33;  DB 12;  Length 298;
  Best Local Similarity   87.5%;  Pred. No. 62;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LSGANINL 9
             | ||||||
Db     256 LRGANINL 263


RESULT  13
O84150
ID   O84150      PRELIMINARY;      PRT;   507 AA.
AC   O84150;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MONOOXYGENASE.
```

```
GN   MHPA.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D/UW-3/CX;
RX   MEDLINE; 99000809.
RA   Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA   Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA   Davis R.W.;
RT   "Genome sequence of an obligate intracellular pathogen of humans:
RT   Chlamydia trachomatis.";
RL   Science 282:754-759(1998).
DR   EMBL; AE001288; AAC67739.1; -.
KW   Monooxygenase.
SQ   SEQUENCE   507 AA;  57864 MW;  2361E01C40A00D13 CRC64;

  Query Match             73.3%;  Score 33;  DB 2;  Length 507;
  Best Local Similarity   66.7%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
             |||| | |:
Db     290 YLSGVNSNI 298


RESULT  14
P94146
ID   P94146      PRELIMINARY;      PRT;   635 AA.
AC   P94146;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   POLY(3-HYDROXYBUTYRATE) DEPOLYMERASE.
GN   PHAZ.
OS   Alcaligenes faecalis.
OC   Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC   Alcaligenes.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AE122;
RA   Kita K., Mashiba S., Ishimaru K., Yanase H., Kato N.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U55775; AAB40611.1; -.
DR   INTERPRO; IPR000379; -.
SQ   SEQUENCE   635 AA;  65212 MW;  CAF65ACA9793C9B9 CRC64;


  Query Match             73.3%;  Score 33;  DB 2;  Length 635;
  Best Local Similarity   75.0%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANIN 8
             |:|||:||
Db     287 YVSGASIN 294


RESULT  15
O26339
ID   O26339      PRELIMINARY;      PRT;  1321 AA.
AC   O26339;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   MAGNESIUM CHELATASE SUBUNIT.
GN   MTH237.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC   Methanobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
```

```
RX   MEDLINE; 98037514.
RA   Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
DR   EMBL; AE000810; AAB84743.1; -.
SQ   SEQUENCE   1321 AA;  144743 MW;  F7AC85F7972143B2 CRC64;


  Query Match             73.3%;  Score 33;  DB 1;  Length 1321;
  Best Local Similarity   75.0%;  Pred. No. 3.3e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANIN 8
            ||:| |||
Db   1081 YLTGRNIN 1088


Search completed: December 16, 2000, 02:00:58
Job time: 5932 sec
```

```
                     GenCore version 4.5
             Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 03:05:20 ; Search time 69.45 Seconds
                                                (without alignments)
                                                4.140 Million cell updates/sec

Title:           US-09-529-121-4
Perfect score:   45
Sequence:        1 YLSGANINL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters:      87993

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_39:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                  %
Result           Query
  No.    Score   Match Length DB  ID                         Description
-------------------------------------------------------------------------------
    1      43    95.6    702  1  CCEM_HUMAN                 P06731 homo sapien
    2      35    77.8    177  1  IL1X_HUMAN                 P18510 homo sapien
    3      34    75.6    335  1  YJ89_PYRHO                 O57713 pyrococcus
    4      34    75.6    956  1  GCSP_ECOLI                 P33195 escherichia
    5      34    75.6   1004  1  GCSP_CHICK                 P15505 gallus gall
    6      34    75.6   1020  1  GCSP_HUMAN                 P23378 homo sapien
    7      34    75.6   1034  1  GCSP_YEAST                 P49095 saccharomyc
    8      33    73.3    514  1  THD1_ECOLI                 P04968 escherichia
    9      33    73.3    514  1  THD1_SALTY                 P20506 salmonella
   10      33    73.3    741  1  RN5A_HUMAN                 Q05823 homo sapien
   11      33    73.3   1286  1  AIDA_ECOLI                 Q03155 escherichia
   12      32    71.1    455  1  NTPA_PEA                   P52914 pisum sativ
   13      32    71.1    510  1  YDYD_SCHPO                 O13692 schizosacch
   14      32    71.1    513  1  THD1_HAEIN                 P46493 haemophilus
   15      32    71.1    550  1  GLYP_CHICK                 P50593 gallus gall
   16      32    71.1    941  1  GCSP_MYCTU                 Q50601 mycobacteri
   17      32    71.1   1034  1  GCSB_FLAPR                 P49362 flaveria pr
   18      32    71.1   1034  1  GCSP_FLAAN                 O49850 flaveria an
   19      32    71.1   1034  1  GCSP_FLATR                 O49852 flaveria tr
   20      32    71.1   1035  1  GCSP_SOLTU                 O49954 solanum tub
   21      32    71.1   1037  1  GCSA_FLAPR                 P49361 flaveria pr
   22      32    71.1   1057  1  GCSP_PEA                   P26969 pisum sativ
   23      32    71.1   3587  1  TYCB_BACBR                 O30408 b tyrocidin
   24      32    71.1   4451  1  GRSB_BACBR                 P14688 b gramicidi
   25      31    68.9    177  1  TRF4_ECOLI                 Q03450 escherichia
   26      31    68.9    248  1  PSPA_MOUSE                 P35242 mus musculu
   27      31    68.9    248  1  PSPA_RAT                   P08427 rattus norv
   28      31    68.9    263  1  AROE_HELPY                 P56119 helicobacte
   29      31    68.9    266  1  AROE_HELPJ                 Q9zjx8 helicobacte
   30      31    68.9    313  1  SURA_HAEIN                 P44721 haemophilus
   31      31    68.9    363  1  V363_ASFB7                 P23164 african swi
   32      31    68.9    455  1  PEX3_PICPA                 Q92262 pichia past
   33      31    68.9    484  1  PER2_VOLCA                 P81132 volvox cart
   34      31    68.9    576  1  THDH_YEAST                 P00927 saccharomyc
   35      31    68.9    623  1  Y014_MYCGE                 P47260 mycoplasma
   36      31    68.9    679  1  RN5A_MOUSE                 Q05921 mus musculu
   37      31    68.9    769  1  YCYL_CAEEL                 Q22523 caenorhabdi
   38      31    68.9    827  1  TRNL_YEAST                 P09880 saccharomyc
   39      31    68.9    933  1  ODO1_ECOLI                 P07015 escherichia
   40    30.5    67.8    377  1  Y412_MYCGE                 P47652 mycoplasma
   41      30    66.7    108  1  HLYU_VIBCH                 P52695 vibrio chol
   42      30    66.7    178  1  IL1X_MOUSE                 P25085 mus musculu
   43      30    66.7    211  1  YI21_BURCE                 P24536 burkholderi
   44      30    66.7    270  1  PMIP_NICAL                 P49173 nicotiana a
   45      30    66.7    350  1  PRIM_METJA                 Q58249 methanococc

                               ALIGNMENTS

RESULT   1
CCEM_HUMAN
ID   CCEM_HUMAN     STANDARD;      PRT;   702 AA.
AC   P06731;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINOEMBRYONIC ANTIGEN PRECURSOR (CEA) (MECONIUM ANTIGEN 100)
DE   (CD66E ANTIGEN).
GN   CEA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90258861.
RA   Schrewe H., Thompson J., Bona M., Hefta L.J.F., Maruya A.,
RA   Hassauer M., Shively J.E., von Kleist S., Zimmermann W.;
RT   "Cloning of the complete gene for carcinoembryonic antigen: analysis
RT   of its promoter indicates a region conveying cell type-specific
RT   expression.";
RL   Mol. Cell. Biol. 10:2738-2748(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88038876.
RA   Beauchemin N., Benchimol S., Cournoyer D., Fuks A., Stanners C.P.;
RT   "Isolation and characterization of full-length functional cDNA clones
RT   for human carcinoembryonic antigen.";
RL   Mol. Cell. Biol. 7:3221-3230(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89122014.
RA   Barnett T., Goebel S.J., Nothdurft M.A., Elting J.J.;
RT   "Carcinoembryonic antigen family: characterization of cDNAs coding
RT   for NCA and CEA and suggestion of nonrandom sequence variation in
RT   their conserved loop-domains.";
RL   Genomics 3:59-66(1988).
RN   [4]
RP   SEQUENCE OF 5-702 FROM N.A.
RX   MEDLINE; 87128144.
RA   Oikawa S., Nakazato H., Kosaki G.;
RT   "Primary structure of human carcinoembryonic antigen (CEA) deduced
RT   from cDNA sequence.";
RL   Biochem. Biophys. Res. Commun. 142:511-518(1987).
RN   [5]
RP   SEQUENCE OF 331-702 FROM N.A.
RX   MEDLINE; 87204247.
RA   Zimmermann W., Ortlieb B., Friedrich R., von Kleist S.;
RT   "Isolation and characterization of cDNA clones encoding the human
RT   carcinoembryonic antigen reveal a highly conserved repeating
RT   structure.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:2960-2964(1987).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC   -!- TISSUE SPECIFICITY: FOUND IN ADENOCARCINOMAS OF ENDODERMALLY
CC       DERIVED DIGESTIVE SYSTEM EPITHELIUM AND FETAL COLON.
CC   -!- PTM: COMPLEX IMMUNOREACTIVE GLYCOPROTEIN WITH A MW OF 180 KDA
```

```
CC       COMPRISING 60% CARBOHYDRATE.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       7 IG-LIKE DOMAINS. BELONGS TO THE CARCINOEMBRYONIC ANTIGEN
CC       SUBFAMILY.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD66e entry;
CC       WWW="HTTP://WWW.NCBI.NLM.NIH.GOV/PROW/CD/CD66E.HTM".
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M17303; AAB59513.1; -.
DR   EMBL; M59262; AAA62835.1; ALT_SEQ.
DR   EMBL; M59255; AAA62835.1; JOINED.
DR   EMBL; M59256; AAA62835.1; JOINED.
DR   EMBL; M59257; AAA62835.1; JOINED.
DR   EMBL; M59258; AAA62835.1; JOINED.
DR   EMBL; M59259; AAA62835.1; JOINED.
DR   EMBL; M59260; AAA62835.1; JOINED.
DR   EMBL; M59261; AAA62835.1; JOINED.
DR   EMBL; M59709; -; NOT_ANNOTATED_CDS.
DR   EMBL; M59710; -; NOT_ANNOTATED_CDS.
DR   EMBL; M29540; AAA51967.1; -.
DR   EMBL; X16455; CAA34474.1; -.
DR   EMBL; M15042; AAA51963.1; -.
DR   EMBL; M16234; AAA51972.1; -.
DR   PIR; A36319; A36319.
DR   MIM; 114890; -.
DR   INTERPRO; IPR003006; -.
DR   PFAM; PF00047; ig; 7.
KW   Immunoglobulin domain; Glycoprotein; GPI-anchor; Membrane; Signal.
FT   SIGNAL        1     34
FT   CHAIN        35      ?       CARCINOEMBRYONIC ANTIGEN.
FT   PROPEP        ?    702       REMOVED IN MATURE FORM.
FT   DOMAIN       35    144       IG-LIKE DOMAIN 1.
FT   DOMAIN      146    237       IG-LIKE DOMAIN 2.
FT   DOMAIN      238    322       IG-LIKE DOMAIN 3.
FT   DOMAIN      324    415       IG-LIKE DOMAIN 4.
FT   DOMAIN      416    498       IG-LIKE DOMAIN 5.
FT   DOMAIN      502    593       IG-LIKE DOMAIN 6.
FT   DOMAIN      594    677       IG-LIKE DOMAIN 7.
FT   CARBOHYD    104    104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    115    115       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    152    152       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    182    182       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    197    197       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    204    204       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    208    208       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    246    246       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    256    256       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    274    274       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    288    288       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    292    292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    309    309       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    330    330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    351    351       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    360    360       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    375    375       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    432    432       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    466    466       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    480    480       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    508    508       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    529    529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    553    553       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    560    560       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    580    580       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    612    612       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    650    650       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    665    665       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CONFLICT    320    320       MISSING (IN REF. 4).
SQ   SEQUENCE   702 AA;  76795 MW;  6299AE26CDDBDB5C CRC64;


  Query Match            95.6%;  Score 43;  DB 1;  Length 702;
  Best Local Similarity   88.9%;  Pred. No. 0.33;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
           ||||||:||
Db     605 YLSGANLNL 613


RESULT   2
IL1X_HUMAN
ID   IL1X_HUMAN     STANDARD;      PRT;   177 AA.
AC   P18510;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   INTERLEUKIN-1 RECEPTOR ANTAGONIST PROTEIN PRECURSOR (IL-1RA) (ICIL-
DE   1RA) (IRAP) (IL-1RN).
GN   IL1RN OR IL1RA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90220867.
RA   Carter D.B., Deibel M.R. Jr., Dunn C.J., Tomich C.S.C., Laborde A.L.,
RA   Slightom J.L., Berger A.E., Bienkowski M.J., Sun F.F., McEwan R.N.,
RA   Harris P.K.W., Yem A.W., Waszak G.A., Chosay J.G., Sieu L.C.,
RA   Hardee M.M., Zurcher-Neely H.A., Reardon I.M., Heinrikson R.L.,
RA   Truesdell S.E., Shelly J.A., Eessalu T.E., Taylor B.M., Tracey D.E.;
RT   "Purification, cloning, expression and biological characterization of
RT   an interleukin-1 receptor antagonist protein.";
RL   Nature 344:633-638(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90136921.
RA   Eisenberg S.P., Evans R.J., Arend W.P., Verderber E., Brewer M.T.,
RA   Hannum C.H., Thompson R.C.;
RT   "Primary structure and functional expression from complementary DNA
RT   of a human interleukin-1 receptor antagonist.";
RL   Nature 343:341-346(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91271363.
RA   Eisenberg S.P., Brewer M.T., Verderber E., Heimdal P.,
RA   Brandhuber B.J., Thompson R.C.;
RT   "Interleukin 1 receptor antagonist is a member of the interleukin 1
RT   gene family: evolution of a cytokine control mechanism.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:5232-5236(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92338323.
RA   Lennard A., Gorman P., Carrier M., Griffiths S., Scotney H.,
RA   Sheer D., Solari R.;
RT   "Cloning and chromosome mapping of the human interleukin-1 receptor
RT   antagonist gene.";
RL   Cytokine 4:83-89(1992).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97146044.
RA   Jenkins J.K., Drong R.F., Shuck M.E., Bienkowski M.J., Slightom J.L.,
RA   Arend W.P., Smith M.F. Jr.;
RT   "Intracellular IL-1 receptor antagonist promoter: cell type-specific
RT   and inducible regulatory regions.";
RL   J. Immunol. 158:748-755(1997).
RN   [6]
RP   SEQUENCE OF 26-45.
RX   MEDLINE; 90136920.
RA   Hannum C.H., Wilcox C.J., Arend W.P., Joslin F.G., Dripps D.J.,
```

```
RA   Heimdal P.L., Armes L.G., Sommer A., Eisenberg S.P., Thompson R.C.;
RT   "Interleukin-1 receptor antagonist activity of a human interleukin-1
RT   inhibitor.";
RL   Nature 343:336-340(1990).
RN   [7]
RP   SEQUENCE OF 26-52.
RX   MEDLINE; 90354444.
RA   Bienkowski M.J., Eessalu T.E., Berger A.E., Truesdell S.E.,
RA   Shelly J.A., Laborde A.L., Zurcher-Neely H.A., Reardon I.M.,
RA   Heinrikson R.L., Chosay J.G., Tracey D.E.;
RT   "Purification and characterization of interleukin 1 receptor level
RT   antagonist proteins from THP-1 cells.";
RL   J. Biol. Chem. 265:14505-14511(1990).
RN   [8]
RP   SEQUENCE FROM N.A. (INTRACELLULAR FORM).
RX   MEDLINE; 91219436.
RA   Haskill S., Martin G., van Le L., Morris J., Peace A., Bigler C.F.,
RA   Jaffe G.J., Hammerberg C., Sporn S.A., Fong S., Arend W.P., Ralph P.;
RT   "cDNA cloning of an intracellular form of the human interleukin 1
RT   receptor antagonist associated with epithelium.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:3681-3685(1991).
RN   [9]
RP   STRUCTURE BY NMR.
RX   MEDLINE; 92297633.
RA   Stockman B.J., Scahill T.A., Roy M., Ulrich E.L., Strakalaitis N.A.,
RA   Brunner D.P., Yem A.W., Deibel M.R. Jr.;
RT   "Secondary structure and topology of interleukin-1 receptor
RT   antagonist protein determined by heteronuclear three-dimensional NMR
RT   spectroscopy.";
RL   Biochemistry 31:5237-5244(1992).
RN   [10]
RP   STRUCTURE BY NMR.
RX   MEDLINE; 94320651.
RA   Stockman B.J., Scahill T.A., Strakalaitis N.A., Brunner D.P.,
RA   Yem A.W., Deibel M.R. Jr.;
RT   "Solution structure of human interleukin-1 receptor antagonist
RT   protein.";
RL   FEBS Lett. 349:79-83(1994).
RN   [11]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX   MEDLINE; 94230368.
RA   Vigers G.P.A., Caffes P., Evans R.J., Thompson R.C., Eisenberg S.P.,
RA   Brandhuber B.J.;
RT   "X-ray structure of interleukin-1 receptor antagonist at 2.0-A
RT   resolution.";
RL   J. Biol. Chem. 269:12874-12879(1994).
RN   [12]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX   MEDLINE; 95172072.
RA   Schreuder H.A., Rondeau J.-M., Tardif C., Soffientini A., Sarubbi E.,
RA   Akeson A., Bowlin T.L., Yanofsky S., Barrett R.W.;
RT   "Refined crystal structure of the interleukin-1 receptor antagonist.
RT   Presence of a disulfide link and a cis-proline.";
RL   Eur. J. Biochem. 227:838-847(1995).
RN   [13]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 32-177 IN COMPLEX WITH IL1R.
RX   MEDLINE; 97215904.
RA   Schreuder H., Tardif C., Trump-Kallmeyer S., Soffientini A.,
RA   Sarubbi E., Akeson A., Bowlin T., Yanofsky S., Barrett R.W.;
RT   "A new cytokine-receptor binding mode revealed by the crystal
RT   structure of the IL-1 receptor with an antagonist.";
RL   Nature 386:194-200(1997).
CC   -!- FUNCTION: IL-1RA INHIBITS THE ACTIVITY OF IL-1 BY BINDING TO ITS
CC       RECEPTOR. IL-1RA HAS NO IL-1 LIKE ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: SECRETED OR INTRACELLULAR (THE VARIANT
CC       FORM).
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE
CC       SPLICING.
CC   -!- TISSUE SPECIFICITY: THE INTRACELLULAR FORM OF IL-1RA IS
CC       PREDOMINANTLY EXPRESSED IN EPITHELIAL CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IL-1 FAMILY.
CC   -!- DATABASE: NAME=R&D Systems' cytokine source book;
CC       WWW="HTTP://WWW.RNDSYSTEMS.COM/CYT_CAT/IL1RA.HTML".
```

```
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M55646; AAA59138.1; -.
DR   EMBL; M63099; AAB41943.1; -.
DR   EMBL; X52015; CAA36262.1; -.
DR   EMBL; X53296; CAA37386.1; -.
DR   EMBL; X64532; CAA45832.1; -.
DR   EMBL; U65590; AAB92268.1; -.
DR   EMBL; U65590; AAB92270.1; -.
DR   PIR; A30368; A30368.
DR   PIR; A37822; A37822.
DR   PIR; S08160; S08160.
DR   PIR; S08159; S08159.
DR   PIR; A40956; A40956.
DR   PIR; A39386; A39386.
DR   PDB; 1ITN; 30-APR-94.
DR   PDB; 2IRT; 15-OCT-94.
DR   PDB; 1IRP; 27-FEB-95.
DR   PDB; 1ILR; 07-FEB-95.
DR   PDB; 1ILT; 01-APR-95.
DR   PDB; 1IRA; 17-JUN-98.
DR   AARHUS/GHENT-2DPAGE; 7104; IEF.
DR   AARHUS/GHENT-2DPAGE; 7105; IEF.
DR   MIM; 147679; -.
DR   INTERPRO; IPR000975; -.
DR   PFAM; PF00340; interleukin-1; 1.
DR   PRINTS; PR00264; INTERLEUKIN1.
DR   PROSITE; PS00253; INTERLEUKIN_1; 1.
KW   Glycoprotein; Signal; Alternative splicing; 3D-structure.
FT   SIGNAL        1     25
FT   CHAIN        26    177       INTERLEUKIN-1 RECEPTOR ANTAGONIST
FT                                PROTEIN.
FT   DISULFID     91    141
FT   CARBOHYD    109    109       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC      1     21       MEICRGLRSHLITLLLFLFHS -> MAL (IN
FT                                INTRACELLULAR ISOFORM).
SQ   SEQUENCE   177 AA;  20055 MW;  D1690776A7394057 CRC64;


  Query Match             77.8%;  Score 35;  DB 1;  Length 177;
  Best Local Similarity   66.7%;  Pred. No. 3.4;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 YLSGANINL 9
          || | |:||
Db     59 YLQGPNVNL 67


RESULT   3
YJ89_PYRHO
ID   YJ89_PYRHO     STANDARD;      PRT;   335 AA.
AC   O57713;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PROTEIN PH1989.
GN   PH1989.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE; 98344137.
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
```

```
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE UPF0104 FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AP000007; BAA31116.1; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM      4     24       POTENTIAL.
FT   TRANSMEM     34     54       POTENTIAL.
FT   TRANSMEM     62     82       POTENTIAL.
FT   TRANSMEM    122    142       POTENTIAL.
FT   TRANSMEM    148    168       POTENTIAL.
FT   TRANSMEM    231    251       POTENTIAL.
FT   TRANSMEM    266    286       POTENTIAL.
FT   TRANSMEM    304    324       POTENTIAL.
SQ   SEQUENCE   335 AA;  37502 MW;  95745BC337B7B494 CRC64;


  Query Match             75.6%;  Score 34;  DB 1;  Length 335;
  Best Local Similarity   66.7%;  Pred. No. 11;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGANINL 9
            :| |||||:
Db       59 FLKGANINV 67


RESULT   4
GCSP_ECOLI
ID   GCSP_ECOLI     STANDARD;      PRT;   956 AA.
AC   P33195;
DT   01-OCT-1993 (Rel. 27, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GLYCINE DEHYDROGENASE [DECARBOXYLATING] (EC 1.4.4.2) (GLYCINE
DE   DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-PROTEIN).
GN   GCVP.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=K12;
RX   MEDLINE; 94237484.
RA   Stauffer L.T., Fogarty S.J., Stauffer G.V.;
RT   "Characterization of the Escherichia coli gcv operon.";
RL   Gene 142:17-22(1994).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC   STRAIN=K12 / W3110;
RX   MEDLINE; 93387305.
RA   Okamura-Ikeda K., Ohmura Y., Fujiwara K., Motokawa Y.;
RT   "Cloning and nucleotide sequence of the gcv operon encoding the
RT   Escherichia coli glycine-cleavage system.";
RL   Eur. J. Biochem. 216:539-548(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
```

```
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC       GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE
CC       THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND
CC       THE REMAINING METHYLAMINE MOIETY IS THE THEN TRANSFERRED TO THE
CC       LIPOAMIDE COFACTOR OF THE H PROTEIN.
CC   -!- CATALYTIC ACTIVITY: GLYCINE + LIPOYLPROTEIN = S-AMINOMETHYL-
CC       DIHYDROLIPOYLPROTEIN + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- SUBUNIT: THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF FOUR PROTEINS:
CC       P, T, L, AND H.
CC   -!- INDUCTION: BY GLYCINE.
CC   -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L20872; AAA23867.1; -.
DR   EMBL; X73958; CAA52146.1; -.
DR   EMBL; U28377; AAA69071.1; -.
DR   EMBL; AE000373; AAC75941.1; -.
DR   PIR; S36834; S36834.
DR   ECOGENE; EG11810; GCVP.
KW   Oxidoreductase; Pyridoxal phosphate.
FT   INIT_MET      0      0
FT   BINDING     707    707       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   956 AA;  104245 MW;  28B7A78D1D0DB6F7 CRC64;


  Query Match             75.6%;  Score 34;  DB 1;  Length 956;
  Best Local Similarity   75.0%;  Pred. No. 32;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGANIN 8
            || |||:|
Db      678 YLDGANMN 685


RESULT   5
GCSP_CHICK
ID   GCSP_CHICK     STANDARD;      PRT;  1004 AA.
AC   P15505;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLYCINE DEHYDROGENASE [DECARBOXYLATING], MITOCHONDRIAL PRECURSOR
DE   (EC 1.4.4.2) (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-
DE   PROTEIN).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91131643.
RA   Kume A., Koyata H., Sakakibara T., Ishiguro Y., Kure S., Hiraga K.;
RT   "The glycine cleavage system. Molecular cloning of the chicken and
RT   human glycine decarboxylase cDNAs and some characteristics involved
RT   in the deduced protein structures.";
RL   J. Biol. Chem. 266:3323-3329(1991).
RN   [2]
RP   SEQUENCE OF 704-757.
RC   TISSUE=LIVER;
RX   MEDLINE; 88106483.
RA   Fujiwara K., Okamura-Ikeda K., Motokawa Y.;
```

```
RT   "Amino acid sequence of the phosphopyridoxyl peptide from P-protein
RT   of the chicken liver glycine cleavage system.";
RL   Biochem. Biophys. Res. Commun. 149:621-627(1987).
CC   -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC       GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE
CC       THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND
CC       THE REMAINING METHYLAMINE MOIETY IS THE THEN TRANSFERRED TO THE
CC       LIPOAMIDE COFACTOR OF THE H PROTEIN.
CC   -!- CATALYTIC ACTIVITY: GLYCINE + LIPOYLPROTEIN = S-AMINOMETHYL-
CC       DIHYDROLIPOYLPROTEIN + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- SUBUNIT: HOMODIMER. THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF
CC       FOUR PROTEINS: P, T, L, AND H.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC   -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M64402; AAA49029.1; -.
DR   EMBL; D90266; BAA14313.1; -.
DR   EMBL; D90240; BAA14287.1; -.
DR   PIR; A27483; A27483.
DR   PIR; A39521; A39521.
KW   Oxidoreductase; Mitochondrion; Pyridoxal phosphate; Transit peptide.
FT   TRANSIT       1      ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN         ?   1004       GLYCINE DEHYDROGENASE [DECARBOXYLATING].
FT   BINDING     738    738       PYRIDOXAL PHOSPHATE.
SQ   SEQUENCE   1004 AA;  111852 MW;  4446D7C66E0DC4BD CRC64;


  Query Match             75.6%;  Score 34;  DB 1;  Length 1004;
  Best Local Similarity   75.0%;  Pred. No. 34;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANIN 8
           || |||:|
Db     709 YLDGANMN 716


RESULT   6
GCSP_HUMAN
ID   GCSP_HUMAN     STANDARD;      PRT;  1020 AA.
AC   P23378;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   GLYCINE DEHYDROGENASE [DECARBOXYLATING], MITOCHONDRIAL PRECURSOR
DE   (EC 1.4.4.2) (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-
DE   PROTEIN).
GN   GCSP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91144593.
RA   Kure S., Narisawa K., Tada K.;
RT   "Structural and expression analyses of normal and mutant mRNA
RT   encoding glycine decarboxylase: three-base deletion in mRNA causes
RT   nonketotic hyperglycinemia.";
RL   Biochem. Biophys. Res. Commun. 174:1176-1182(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91131643.
RA   Kume A., Koyata H., Sakakibara T., Ishiguro Y., Kure S., Hiraga K.;
RT   "The glycine cleavage system. Molecular cloning of the chicken and
RT   human glycine decarboxylase cDNAs and some characteristics involved
RT   in the deduced protein structures.";
RL   J. Biol. Chem. 266:3323-3329(1991).
RN   [3]
RP   VARIANT NKH1 ILE-564.
RX   MEDLINE; 92340654.
RA   Kure S., Takayanagi M., Narisawa K., Tada K., Leisti J.;
RT   "Identification of a common mutation in Finnish patients with
RT   nonketotic hyperglycinemia.";
RL   J. Clin. Invest. 90:160-164(1992).
CC   -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC       GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE
CC       THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND
CC       THE REMAINING METHYLAMINE MOIETY IS THE THEN TRANSFERRED TO THE
CC       LIPOAMIDE COFACTOR OF THE H PROTEIN.
CC   -!- CATALYTIC ACTIVITY: GLYCINE + LIPOYLPROTEIN = S-AMINOMETHYL-
CC       DIHYDROLIPOYLPROTEIN + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- SUBUNIT: HOMODIMER. THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF
CC       FOUR PROTEINS: P, T, L, AND H.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC   -!- DISEASE: DEFECTS IN GCSP ARE THE CAUSE OF TYPE I NONKETOTIC
CC       HYPERGLYCINEMIA (NKH1). NKH IS AN AUTOSOMAL RECESSIVE DISEASE
CC       CHARACTERIZED BY ACCUMULATION OF A LARGE AMOUNT OF GLYCINE IN BODY
CC       FLUID AND BY SEVERE NEUROLOGICAL SYMPTOMES.
CC   -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M63635; AAA36478.1; -.
DR   EMBL; M64590; AAA36463.1; -.
DR   EMBL; D90239; BAA14286.1; -.
DR   PIR; B39521; B39521.
DR   PIR; JN0124; JN0124.
DR   MIM; 238300; -.
KW   Oxidoreductase; Mitochondrion; Pyridoxal phosphate; Transit peptide;
KW   Disease mutation.
FT   TRANSIT       1     35       MITOCHONDRION (POTENTIAL).
FT   CHAIN        36   1020       GLYCINE DEHYDROGENASE [DECARBOXYLATING].
FT   BINDING     754    754       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   VARIANT     564    564       S -> I (IN NKH1).
FT                                /FTId=VAR_004979.
FT   CONFLICT    396    396       A -> R (IN REF. 2).
FT   CONFLICT    608    608       H -> Y (IN REF. 2).
FT   CONFLICT    976    976       V -> M (IN REF. 2).
SQ   SEQUENCE   1020 AA;  112712 MW;  129FAF26B8D37E0F CRC64;


  Query Match             75.6%;  Score 34;  DB 1;  Length 1020;
  Best Local Similarity   75.0%;  Pred. No. 35;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANIN 8
           || |||:|
Db     725 YLDGANMN 732


RESULT   7
GCSP_YEAST
ID   GCSP_YEAST     STANDARD;      PRT;  1034 AA.
AC   P49095;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLYCINE DEHYDROGENASE [DECARBOXYLATING], MITOCHONDRIAL PRECURSOR
DE   (EC 1.4.4.2) (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-
DE   PROTEIN).
GN   GCV2 OR GSD2 OR YMR189W OR YM9646.01.
```

```
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96120340.
RA   Sinclair D.A., Dawes I.W.;
RT   "Genetics of the synthesis of serine from glycine and the utilization
RT   of glycine as sole nitrogen source by Saccharomyces cerevisiae.";
RL   Genetics 140:1213-1222(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Pearson D., Bowman S., Barrell B.G., Rajandream M.A.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC       GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE
CC       THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND
CC       THE REMAINING METHYLAMINE MOIETY IS THE THEN TRANSFERRED TO THE
CC       LIPOAMIDE COFACTOR OF THE H PROTEIN.
CC   -!- CATALYTIC ACTIVITY: GLYCINE + LIPOYLPROTEIN = S-AMINOMETHYL-
CC       DIHYDROLIPOYLPROTEIN + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL (BY SIMILARITY).
CC   -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U20641; AAB18933.1; -.
DR   EMBL; Z47815; CAA87810.1; -.
DR   SGD; S0004801; GCV2.
KW   Oxidoreductase; Mitochondrion; Pyridoxal phosphate; Transit peptide.
FT   TRANSIT       1      ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN         ?   1034       GLYCINE DEHYDROGENASE [DECARBOXYLATING].
FT   BINDING     773    773       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   1034 AA;  114451 MW;  F4D52642B0BDA041 CRC64;


  Query Match             75.6%;  Score 34;  DB 1;  Length 1034;
  Best Local Similarity   75.0%;  Pred. No. 35;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANIN 8
             || |||:|
Db     744 YLDGANMN 751


RESULT   8
THD1_ECOLI
ID   THD1_ECOLI     STANDARD;      PRT;   514 AA.
AC   P04968;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   THREONINE DEHYDRATASE BIOSYNTHETIC (EC 4.2.1.16) (THREONINE
DE   DEAMINASE).
GN   ILVA.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 88056322.
RA   Cox J.L., Cox B.J., Fidanza V., Calhoun D.H.;
RT   "The complete nucleotide sequence of the ilvGMEDA cluster of
RT   Escherichia coli K-12.";
RL   Gene 56:185-198(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Garrison E., Harms E., Umbarger H.E.;
RL   Submitted (AUG-1986) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 87174741.
RA   Lawther R.P., Wek R.C., Lopes J.M., Pereira R., Taillon B.E.,
RA   Hatfield G.W.;
RT   "The complete nucleotide sequence of the ilvGMEDA operon of
RT   Escherichia coli K-12.";
RL   Nucleic Acids Res. 15:2137-2155(1987).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 92358234.
RA   Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT   "Analysis of the Escherichia coli genome: DNA sequence of the region
RT   from 84.5 to 86.5 minutes.";
RL   Science 257:771-778(1992).
RN   [5]
RP   SEQUENCE OF 439-514 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 86111952.
RA   Wek R.C., Hatfield G.W.;
RT   "Nucleotide sequence and in vivo expression of the ilvY and ilvC
RT   genes in Escherichia coli K12. Transcription from divergent
RT   overlapping promoters.";
RL   J. Biol. Chem. 261:2441-2450(1986).
RN   [6]
RP   SEQUENCE OF 1-10 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 89326124.
RA   Lopes J.M., Lawther R.P.;
RT   "Physical identification of an internal promoter, ilvAp, in the
RT   distal portion of the ilvGMEDA operon.";
RL   Gene 76:255-269(1989).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX   MEDLINE; 98230745.
RA   Gallagher D.T., Gilliland G.L., Xiao G., Zondlo J., Fisher K.E.,
RA   Chinchilla D., Eisenstein E.;
RT   "Structure and control of pyridoxal phosphate dependent allosteric
RT   threonine deaminase.";
RL   Structure 6:465-475(1998).
CC   -!- FUNCTION: DEAMINATES L-THREONINE, BUT ALSO L-SERINE TO A LESSER
CC       EXTENT.
CC   -!- CATALYTIC ACTIVITY: L-THREONINE + H(2)O = 2-OXOBUTANOATE + NH(3)
CC       + H(2)O.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- ENZYME REGULATION: ISOLEUCINE ALLOSTERICALLY INHIBITS WHILE
CC       VALINE ALLOSTERICALLY ACTIVATES THIS ENZYME.
CC   -!- PATHWAY: FIRST STEP OF ISOLEUCINE BIOSYNTHESIS.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SIMILARITY: TO THREONINE DEHYDRATASE CATABOLIC AND SERINE
CC       DEHYDRATASE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X04890; CAA28577.1; -.
DR   EMBL; K03503; AAA24014.1; -.
DR   EMBL; M10313; AAB59054.1; -.
DR   EMBL; M11689; AAA24027.1; -.
DR   EMBL; M32253; AAA24024.1; -.
DR   EMBL; M87049; AAA67575.1; -.
```

```
DR   EMBL; AE000453; AAC77492.1; -.
DR   EMBL; M25497; AAA24015.1; -.
DR   PIR; B27310; DWECTS.
DR   PIR; S30670; S30670.
DR   PIR; S48895; S48895.
DR   PDB; 1TDJ; 18-NOV-98.
DR   ECO2DBASE; F050.1; 6TH EDITION.
DR   ECOGENE; EG10493; ILVA.
DR   INTERPRO; IPR000634; -.
DR   INTERPRO; IPR001721; -.
DR   INTERPRO; IPR001926; -.
DR   PFAM; PF00291; S_T_dehydratase; 1.
DR   PFAM; PF00585; Thr_dehydrat_C; 2.
DR   PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW   Isoleucine biosynthesis; Lyase; Pyridoxal phosphate;
KW   Allosteric enzyme; 3D-structure.
FT   BINDING      62     62       PYRIDOXAL PHOSPHATE.
FT   CONFLICT    120    120       A -> R (IN REF. 2).
FT   CONFLICT    140    140       A -> R (IN REF. 2).
FT   CONFLICT    195    195       G -> C (IN REF. 2).
FT   CONFLICT    243    243       A -> G (IN REF. 3).
FT   CONFLICT    334    334       G -> V (IN REF. 2).
SQ   SEQUENCE   514 AA;  56195 MW;  9D389A0EDD8DE692 CRC64;


  Query Match             73.3%;  Score 33;  DB 1;  Length 514;
  Best Local Similarity   85.7%;  Pred. No. 27;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSGANIN 8
            |||||:|
Db      314 LSGANVN 320


RESULT   9
THD1_SALTY
ID   THD1_SALTY     STANDARD;      PRT;   514 AA.
AC   P20506;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   THREONINE DEHYDRATASE BIOSYNTHETIC (EC 4.2.1.16) (THREONINE
DE   DEAMINASE).
GN   ILVA.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88255870.
RA   Taillon B.E., Little R., Lawther R.P.;
RT   "Analysis of the functional domains of biosynthetic threonine
RT   deaminase by comparison of the amino acid sequences of three
RT   wild-type alleles to the amino acid sequence of biodegradative
RT   threonine deaminase.";
RL   Gene 63:245-252(1988).
RN   [2]
RP   SEQUENCE OF 1-10 FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE; 89326124.
RA   Lopes J.M., Lawther R.P.;
RT   "Physical identification of an internal promoter, ilvAp, in the
RT   distal portion of the ilvGMEDA operon.";
RL   Gene 76:255-269(1989).
CC   -!- CATALYTIC ACTIVITY: L-THREONINE + H(2)O = 2-OXOBUTANOATE + NH(3)
CC       + H(2)O.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- ENZYME REGULATION: ISOLEUCINE ALLOSTERICALLY INHIBITS WHILE
CC       VALINE ALLOSTERICALLY ACTIVATES THIS ENZYME.
CC   -!- PATHWAY: FIRST STEP OF ISOLEUCINE BIOSYNTHESIS.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SIMILARITY: TO THREONINE DEHYDRATASE CATABOLIC AND SERINE
CC       DEHYDRATASE.
```

```
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M26670; AAA27150.1; -.
DR   EMBL; M25498; AAA27151.1; -.
DR   PIR; JT0278; DWEBTT.
DR   STYGENE; SG10179; ILVA.
DR   INTERPRO; IPR000634; -.
DR   INTERPRO; IPR001721; -.
DR   INTERPRO; IPR001926; -.
DR   PFAM; PF00291; S_T_dehydratase; 1.
DR   PFAM; PF00585; Thr_dehydrat_C; 2.
DR   PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW   Isoleucine biosynthesis; Lyase; Pyridoxal phosphate;
KW   Allosteric enzyme.
FT   BINDING      62     62       PYRIDOXAL PHOSPHATE.
SQ   SEQUENCE   514 AA;  56276 MW;  03B881F542E363AF CRC64;


  Query Match             73.3%;  Score 33;  DB 1;  Length 514;
  Best Local Similarity   85.7%;  Pred. No. 27;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSGANIN 8
            |||||:|
Db      314 LSGANVN 320


RESULT  10
RN5A_HUMAN
ID   RN5A_HUMAN     STANDARD;      PRT;   741 AA.
AC   Q05823;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   2-5A-DEPENDENT RIBONUCLEASE (EC 3.1.26.-) (2-5A-DEPENDENT RNAASE)
DE   (RNASE L) (RIBONUCLEASE 4).
GN   RNS4 OR RNASEL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY;
RX   MEDLINE; 93201598.
RA   Zhou A., Hassel B.A., Silverman R.H.;
RT   "Expression cloning of 2-5A-dependent RNAase: a uniquely regulated
RT   mediator of interferon action.";
RL   Cell 72:753-765(1993).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE; 94245737.
RA   Dong B., Xu L., Zhou A., Hassel B.A., Lee X., Torrence P.F.,
RA   Silverman R.H.;
RT   "Intrinsic molecular activities of the interferon-induced
RT   2-5A-dependent RNase.";
RL   J. Biol. Chem. 269:14153-14158(1994).
CC   -!- FUNCTION: MEDIATES THE EFFECTS OF INTERFERON AGAINST
CC       PICORNAVIRUSES. ACTIVE WHEN BOUND TO 2-5A (5'-PHOSPHORYLATED
CC       2',5'-LINKED OLIGOADENYLATES). CAPABLE OF CLEAVING POLY(RU) AND,
CC       TO A LESSER EXTENT, POLY(RA), TO SETS OF DISCRETE PRODUCTS RANGING
CC       FROM BETWEEN 4 AND 22 NUCLEOTIDES IN LENGTH.
CC   -!- CATALYTIC ACTIVITY: CLEAVES 3' OF UPNP DIMERS, WITH PREFERENCE
CC       FOR UU AND UA SEQUENCES.
CC   -!- COFACTOR: OPTIMAL RNA CLEAVAGE RATES REQUIRES THE PRESENCE OF
CC       EITHER MANGANESE OR MAGNESIUM AND ATP.
CC   -!- SUBUNIT: MONOMER.
```

```
CC   -!- INDUCTION: BY INTERFERON.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L10381; AAA18032.1; -.
DR   HSSP; Q00420; 1AWC.
DR   MIM; 180435; -.
DR   INTERPRO; IPR000719; -.
DR   INTERPRO; IPR002110; -.
DR   PFAM; PF00023; ank; 8.
DR   PFAM; PF00069; pkinase; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 6.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Hydrolase; Nuclease; RNA-binding; Endonuclease; Zinc-finger.
FT   DOMAIN      229    242       2-5A BINDING (P-LOOP).
FT   DOMAIN      253    275       2-5A BINDING (P-LOOP).
FT   DOMAIN      365    586       PROTEIN KINASE LIKE.
FT   ZN_FING     395    444       C6-TYPE (POTENTIAL).
SQ   SEQUENCE   741 AA;  83532 MW;  91385EA307E3CE1D CRC64;

  Query Match             73.3%;  Score 33;  DB 1;  Length 741;
  Best Local Similarity   66.7%;  Pred. No. 40;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              |  |||:||
Db        145 YKRGANVNL 153


RESULT  11
AIDA_ECOLI
ID   AIDA_ECOLI     STANDARD;      PRT;  1286 AA.
AC   Q03155;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   ADHESIN AIDA-I PRECURSOR.
GN   AIDA-I.
OS   Escherichia coli.
OG   Plasmid pIB6.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 50-56.
RC   STRAIN=2787 (O126:H27);
RX   MEDLINE; 92326638.
RA   Benz I., Schmidt M.A.;
RT   "AIDA-I, the adhesin involved in diffuse adherence of the
RT   diarrhoeagenic Escherichia coli strain 2787 (O126:H27), is
RT   synthesized via a precursor molecule.";
RL   Mol. Microbiol. 6:1539-1546(1992).
CC   -!- FUNCTION: THIS IS AN ADHESION PROTEIN NECESSARY FOR THE DIFFUSE
CC       ADHERENCE OF DIARRHEA-CAUSING ENTEROPATHOGENIC ESCHERICHIA COLI
CC       TO EPITHELIAL CELLS.
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X65022; CAA46156.1; -.
DR   PIR; S28634; S28634.
KW   Cell adhesion; Signal; Outer membrane; Plasmid.
FT   SIGNAL        1     49
FT   CHAIN        50      ?       ADHESIN AIDA-I.
FT   PROPEP        ?   1286
SQ   SEQUENCE   1286 AA;  132271 MW;  B2A00F72AC05FB34 CRC64;

  Query Match             73.3%;  Score 33;  DB 1;  Length 1286;
  Best Local Similarity   85.7%;  Pred. No. 71;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 SGANINL 9
              ||||:||
Db        507 SGANVNL 513


RESULT  12
NTPA_PEA
ID   NTPA_PEA       STANDARD;      PRT;   455 AA.
AC   P52914;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   NUCLEOSIDE-TRIPHOSPHATASE (EC 3.6.1.15) (NUCLEOSIDE TRIPHOSPHATE
DE   PHOSPHOHYDROLASE) (NTPASE).
OS   Pisum sativum (Garden pea).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Pisum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. ALASKA; TISSUE=PLUMULE;
RX   MEDLINE; 96197404.
RA   Hsieh H., Tong C.G., Thomas C., Roux S.J.;
RT   "Light-modulated abundance of an mRNA encoding a
RT   calmodulin-regulated, chromatin-associated NTPase in pea.";
RL   Plant Mol. Biol. 30:135-147(1996).
CC   -!- FUNCTION: MIGHT BE INVOLVED IN RNA TRANSPORT OUT OF NUCLEI.
CC   -!- CATALYTIC ACTIVITY: NTP + H(2)O = NDP + PHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR; NUCLEUS.
CC   -!- SIMILARITY: BELONGS TO THE GDA1 / CD39 NTPASE FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z32743; CAA83655.1; -.
DR   INTERPRO; IPR000407; -.
DR   PFAM; PF01150; GDA1_CD39; 1.
DR   PROSITE; PS01238; GDA1_CD39_NTPASE; 1.
KW   Hydrolase; Nuclear protein.
SQ   SEQUENCE   455 AA;  50072 MW;  50FDF0023ABC4299 CRC64;

  Query Match             71.1%;  Score 32;  DB 1;  Length 455;
  Best Local Similarity   75.0%;  Pred. No. 38;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGANIN 8
              | |||| |
Db        287 YTSGANFN 294


RESULT  13
YDYD_SCHPO
ID   YDYD_SCHPO     STANDARD;      PRT;   510 AA.
AC   O13692;
DT   15-JUL-1998 (Rel. 36, Created)
```

```
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   HYPOTHETICAL 53.7 KDA PROTEIN C11E3.13C IN CHROMOSOME I PRECURSOR.
GN   SPAC11E3.13C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC   Schizosaccharomycetaceae; Schizosaccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Murphy L., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE GAS1 FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z98595; CAB11192.1; -.
KW   Hypothetical protein; Glycoprotein; Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    510       HYPOTHETICAL PROTEIN C11E3.13C.
FT   DOMAIN      423    488       SER-RICH.
FT   CARBOHYD    147    147       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    216    216       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    252    252       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    318    318       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    337    337       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    397    397       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    491    491       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   510 AA;  53692 MW;  3D515CCDB5651087 CRC64;


  Query Match             71.1%;  Score 32;  DB 1;  Length 510;
  Best Local Similarity   75.0%;  Pred. No. 43;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSGANIN 8
            || | |||
Db       76 YLQGLNIN 83


RESULT  14
THD1_HAEIN
ID   THD1_HAEIN     STANDARD;      PRT;   513 AA.
AC   P46493;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   THREONINE DEHYDRATASE BIOSYNTHETIC (EC 4.2.1.16) (THREONINE
DE   DEAMINASE).
GN   ILVA OR HI0738.1.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   IDENTIFICATION.
RA   Koonin E.V., Rudd K.E.;
RL   Submitted (SEP-1995) to the SWISS-PROT data bank.
CC   -!- FUNCTION: DEAMINATES L-THREONINE, BUT ALSO L-SERINE TO A LESSER
CC       EXTENT (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: L-THREONINE + H(2)O = 2-OXOBUTANOATE + NH(3)
CC       + H(2)O.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC   -!- ENZYME REGULATION: ISOLEUCINE ALLOSTERICALLY INHIBITS WHILE
CC       VALINE ALLOSTERICALLY ACTIVATES THIS ENZYME (BY SIMILARITY).
CC   -!- PATHWAY: FIRST STEP OF ISOLEUCINE BIOSYNTHESIS.
CC   -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC   -!- SIMILARITY: TO THREONINE DEHYDRATASE CATABOLIC AND SERINE
CC       DEHYDRATASE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U32757; AAC22398.1; -.
DR   TIGR; HI0738.1; -.
DR   INTERPRO; IPR000634; -.
DR   INTERPRO; IPR001721; -.
DR   INTERPRO; IPR001926; -.
DR   PFAM; PF00291; S_T_dehydratase; 1.
DR   PFAM; PF00585; Thr_dehydrat_C; 2.
DR   PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW   Isoleucine biosynthesis; Lyase; Pyridoxal phosphate;
KW   Allosteric enzyme.
FT   BINDING      63     63       PYRIDOXAL PHOSPHATE.
SQ   SEQUENCE   513 AA;  56662 MW;  DF42CA8B6FDE4CD7 CRC64;


  Query Match             71.1%;  Score 32;  DB 1;  Length 513;
  Best Local Similarity   85.7%;  Pred. No. 44;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSGANIN 8
              |||||:|
Db      315 LSGANLN 321


RESULT  15
GLYP_CHICK
ID   GLYP_CHICK     STANDARD;      PRT;   550 AA.
AC   P50593;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   GLYPICAN-1 PRECURSOR (HEPARAN SULFATE PROTEOGLYCAN CORE PROTEIN).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RA   Shi N., Antin P., Flink I.L., Morkin E.;
RL   Submitted (MAY-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CELL SURFACE PROTEOGLYCAN THAT BEARS HEPARAN SULFATE.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE GLYPICAN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L29089; AAA65199.1; ALT_INIT.
DR   INTERPRO; IPR001863; -.
DR   PFAM; PF01153; Glypican; 1.
DR   PROSITE; PS01207; GLYPICAN; 1.
KW   Proteoglycan; Heparan sulfate; Glycoprotein; Signal; GPI-anchor;
KW   Extracellular matrix.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21      ?       GLYPICAN-1.
FT   PROPEP        ?    550       REMOVED IN MATURE FORM (POTENTIAL).
FT   CARBOHYD     76     76       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    113    113       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    382    382       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     52     52       O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT   CARBOHYD    483    483       O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT   CARBOHYD    485    485       O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT   CARBOHYD    487    487       O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61084 MW;  6B7B796E506BF7FF CRC64;


  Query Match             71.1%;  Score 32;  DB 1;  Length 550;
  Best Local Similarity   66.7%;  Pred. No. 47;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 YLSGANINL 9
          |  |:||||
Db    148 YYRGSNINL 156


Search completed: December 16, 2000, 03:05:20
Job time: 8040 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 01:51:15 ; Search time 91.14 Seconds
                                              (without alignments)
                                              6.267 Million cell updates/sec

Title:           US-09-529-121-4
Perfect score:   45
Sequence:        1 YLSGANINL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters:      182106

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_65:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                   %
Result           Query
  No.    Score   Match Length DB  ID                    Description
-------------------------------------------------------------------------
    1       43    95.6    702  2  A36319                carcinoembryonic a
    2       37    82.2    506  2  C81704                monooxygenase-rela
    3       36    80.0    554  2  A70904                probable acid--CoA
    4       35    77.8    177  2  A30368                interleukin-1 rece
    5       35    77.8    180  2  A39386                interleukin-1 rece
    6       34    75.6     83  2  I64001                hypothetical prote
    7       34    75.6    335  2  E71215                hypothetical prote
    8       34    75.6    949  2  E75352                glycine cleavage s
    9       34    75.6    957  2  S36834                glycine dehydrogen
   10       34    75.6    979  2  T16734                hypothetical prote
   11       34    75.6   1003  2  A39521                glycine dehydrogen
   12       34    75.6   1020  2  JN0124                glycine dehydrogen
   13       34    75.6   1020  2  B39521                glycine dehydrogen
   14       34    75.6   1034  2  S50917                aminomethyltransfe
   15       33    73.3    507  2  E71551                probable monooxyge
   16       33    73.3    514  1  DWECTS                threonine dehydrat
   17       33    73.3    514  1  DWEBTT                threonine dehydrat
   18       33    73.3    741  2  A45771                2-5A-dependent RNA
   19       33    73.3    950  2  D81821                glycine dehydrogen
   20       33    73.3   1286  2  S28634                adhesin AIDA-I pre
   21       33    73.3   1321  2  E69129                protoporphyrin IX
   22       33    73.3   1956  2  T16416                hypothetical prote
   23       32    71.1    236  2  T22220                hypothetical prote
   24       32    71.1    248  2  T26461                hypothetical prote
   25       32    71.1    259  2  S76576                hypothetical prote
   26       32    71.1    262  2  C81384                shikimate 5-dehydr
   27       32    71.1    272  2  T22562                hypothetical prote
   28       32    71.1    343  2  T15192                hypothetical prote
   29       32    71.1    356  2  G72386                conserved hypothet
```

```
   30       32    71.1    455  2  S48859                nucleoside triphos
   31       32    71.1    473  2  B81919                probable membrane
   32       32    71.1    473  2  E81182                hypothetical prote
   33       32    71.1    510  2  T37541                probable glycolipi
   34       32    71.1    753  2  T24745                hypothetical prote
   35       32    71.1    941  2  A70722                probable gcvB prot
   36       32    71.1    952  2  T44754                probable glycine d
   37       32    71.1    983  2  S76257                hypothetical prote
   38       32    71.1   1031  2  T46636                glycine dehydrogen
   39       32    71.1   1034  2  S63536                aminomethyltransfe
   40       32    71.1   1035  2  T07826                aminomethyltransfe
   41       32    71.1   1037  2  T05309                probable glycine d
   42       32    71.1   1037  2  S63535                aminomethyltransfe
   43       32    71.1   1037  2  S40216                P protein - Flaver
   44       32    71.1   1044  2  T02615                probable glycine d
   45       32    71.1   1057  2  A42109                glycine dehydrogen

                              ALIGNMENTS

RESULT   1
A36319
carcinoembryonic antigen precursor - human
N;Alternate names: CEA; meconium antigen 100
C;Species: Homo sapiens (man)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 31-Jan-2000
C;Accession: A36319; A27773; A31037; A25845; S08106; S31737; A44476; I54224; I59098;
R;Schrewe, H.; Thompson, J.; Bona, M.; Hefta, L.J.F.; Maruya, A.; Hassauer, M.; Shive
Mol. Cell. Biol. 10, 2738-2748, 1990
A;Title: Cloning of the complete gene for carcinoembryonic antigen: analysis of its p
A;Reference number: A36319; MUID:90258861
A;Accession: A36319
A;Molecule type: DNA
A;Residues: 1-702 <SCH>
A;Cross-references: GB:M17303; NID:g178676; PIDN:AAB59513.1; PID:g178677
A;Note: the authors show the codons TTA for residue 641-Phe and CAG for residue 646-T
R;Beauchemin, N.; Benchimol, S.; Cournoyer, D.; Fuks, A.; Stanners, C.P.
Mol. Cell. Biol. 7, 3221-3230, 1987
A;Title: Isolation and characterization of full-length functional cDNA clones for hum
A;Reference number: A27773; MUID:88038876
A;Accession: A27773
A;Molecule type: mRNA
A;Residues: 1-702 <BEA>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
R;Barnett, T.; Goebel, S.J.; Nothdurft, M.A.; Elting, J.J.
Genomics 3, 59-66, 1988
A;Title: Carcinoembryonic antigen family: characterization of cDNAs coding for NCA an
A;Reference number: A31037; MUID:89122014
A;Accession: A31037
A;Molecule type: mRNA
A;Residues: 1-702 <BAR>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
A;Note: the authors translated the codon GTG for residue 130 as Leu
R;Oikawa, S.; Nakazato, H.; Kosaki, G.
Biochem. Biophys. Res. Commun. 142, 511-518, 1987
A;Title: Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA
A;Reference number: A25845; MUID:87128144
A;Accession: A25845
A;Molecule type: mRNA
A;Residues: 5-702 <OIK>
A;Cross-references: GB:M15042; NID:g180198; PIDN:AAA51963.1; PID:g180199
R;Oikawa, S.
submitted to the EMBL Data Library, September 1989
A;Reference number: S08106
A;Accession: S08106
A;Molecule type: mRNA
A;Residues: 5-319,321-702 <OI2>
A;Cross-references: EMBL:X16455; NID:g29854; PIDN:CAA34474.1; PID:g825638
R;Barnett, T.
submitted to the EMBL Data Library, September 1991
A;Description: Genomic DNA sequence upstream of the translational start of the carcin
A;Reference number: S31737
```

```
A;Accession: S31737
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-141 <BA2>
A;Cross-references: EMBL:X62151
R;Khan, W.N.; Fraengsmyr, L.; Teglund, S.; Israelsson, A.; Bremer, K.; Hammarstroem, S.
Genomics 14, 384-390, 1992
A;Title: Identification of three new genes and estimation of the size of the carcinoembr
A;Reference number: A44476; MUID:93052339
A;Accession: A44476
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 35-141 <KHA>
R;Willcocks, T.C.; Craig, I.W.
Genomics 8, 492-500, 1990
A;Title: Characterization of the genomic organization of human carcinoembryonic antigen
A;Reference number: I54224; MUID:91139118
A;Accession: I54224
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-37 <RES>
A;Cross-references: GB:M60964; NID:g180215; PIDN:AAA51964.1; PID:g180217
R;Zimmermann, W.; Ortlieb, B.; Friedrich, R.; von Kleist, S.
Proc. Natl. Acad. Sci. U.S.A. 84, 2960-2964, 1987
A;Title: Isolation and characterization of cDNA clones encoding the human carcinoembryon
A;Reference number: I59098; MUID:87204247
A;Accession: I59098
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 331-702 <RE2>
A;Cross-references: GB:M16234; NID:g180240; PIDN:AAA51972.1; PID:g180241
R;Siepen, D.; Paxton, R.J.; Neumaier, M.; Shively, J.E.; Wagener, C.
Biochem. Biophys. Res. Commun. 147, 212-218, 1987
A;Title: Carcinoembryonic antigen (CEA) and two crossreacting antigens of 165 KD and 105
A;Reference number: A26831; MUID:87326349
A;Accession: A26831
A;Molecule type: protein
A;Residues: 35-64 <SIE>
R;Thomas, P.; Toth, C.A.
Biochem. Biophys. Res. Commun. 170, 391-396, 1990
A;Title: Carcinoembryonic antigen binding to Kupffer cells is via a peptide located at t
A;Reference number: A35490; MUID:90321257
A;Accession: A35490
A;Molecule type: protein
A;Residues: 'X',140-151,'X',153,'X',155-156 <THO>
A;Note: this is the amino terminal end of a fragment shown to mediate uptake by Kupffer
C;Comment: This heavily glycosylated membane protein of unknown function is a widely use
C;Comment: This protein may be processed at its C-terminus. It is anchored to the membra
C;Genetics:
A;Gene: GDB:CEA
A;Cross-references: GDB:119054; OMIM:114890
A;Map position: 19q13.2-19q13.2
A;Introns: 22/1; 142/1; 235/1; 320/1; 413/1; 498/1; 591/1; 676/1
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin
C;Keywords: blocked carboxyl end; glycoprotein; lipoprotein; membrane protein; phosphati
F;1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-678/Product: carcinoembryonic antigen #status predicted <MAT>
F;160-217/Domain: immunoglobulin homology <IMM1>
F;252-301/Domain: immunoglobulin homology <IMM2>
F;338-395/Domain: immunoglobulin homology <IMM3>
F;516-573/Domain: immunoglobulin homology <IMM4>
F;608-657/Domain: immunoglobulin homology <IMM5>
F;679-702/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F;678/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Gly) (in mature form


  Query Match             95.6%;  Score 43;  DB 2;  Length 702;
  Best Local Similarity   88.9%;  Pred. No. 0.55;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
             ||||||:||
Db     605 YLSGANLNL 613


RESULT   2
C81704
monooxygenase-related protein TC0425 [imported] - Chlamydia muridarum (strain Nigg)
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C;Accession: C81704
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A;Reference number: A81500; MUID:20150255
A;Accession: C81704
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-506 <TET>
A;Cross-references: GB:AE002309; GB:AE002160; NID:g7190464; PIDN:AAF39281.1; PID:g719
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
A;Gene: TC0425


  Query Match             82.2%;  Score 37;  DB 2;  Length 506;
  Best Local Similarity   66.7%;  Pred. No. 7;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
             |||| |:|:
Db     290 YLSGVNLNI 298


RESULT   3
A70904
probable acid--CoA ligase (EC 6.2.1.-) - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70904
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: A70904
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-554 <COL>
A;Cross-references: GB:Z97050; GB:AL123456; NID:g3256008; PIDN:CAB09749.1; PID:g22135
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: fadD5
C;Superfamily: 4-coumarate--CoA ligase; acetate--CoA ligase homology
C;Keywords: acid-thiol ligase
F;72-525/Domain: acetate--CoA ligase homology <ACL>


  Query Match             80.0%;  Score 36;  DB 2;  Length 554;
  Best Local Similarity   87.5%;  Pred. No. 12;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANIN 8
             | ||||||
Db     217 YTSGANIN 224


RESULT   4
A30368
interleukin-1 receptor antagonist secreted form precursor - human
C;Species: Homo sapiens (man)
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 26-May-2000
```

```
C;Accession: A40956; I37894; A30368; S08160; S08159; A37822
R;Eisenberg, S.P.; Brewer, M.T.; Verderber, E.; Heimdal, P.; Brandhuber, B.J.; Thompson,
Proc. Natl. Acad. Sci. U.S.A. 88, 5232-5236, 1991
A;Title: Interleukin 1 receptor antagonist is a member of the interleukin 1 gene family:
A;Reference number: A40956; MUID:91271363
A;Accession: A40956
A;Molecule type: DNA
A;Residues: 1-177 <EIS>
A;Cross-references: GB:M63099; NID:g186385; PIDN:AAB41943.1; PID:g186386
R;Lennard, A.; Gorman, P.; Carrier, M.; Griffiths, S.; Scotney, H.; Sheer, D.; Solari, R
Cytokine 4, 83-89, 1992
A;Title: Cloning and chromosome mapping of the human interleukin-1 receptor antagonist g
A;Reference number: I37894; MUID:92338323
A;Accession: I37894
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-177 <LEN>
A;Cross-references: EMBL:X64532; NID:g33798; PIDN:CAA45832.1; PID:g33799
R;Carter, D.B.; Deibel Jr., M.R.; Dunn, C.J.; Tomich, C.S.C.; Laborde, A.L.; Slightom, J
J.G.; Sieu, L.C.; Hardee, M.M.; Zurcher-Neely, H.A.; Reardon, I.M.; Heinrikson, R.L.; Tr
Nature 344, 633-638, 1990
A;Title: Purification, cloning, expression and biological characterization of an interle
A;Reference number: A30368; MUID:90220867
A;Accession: A30368
A;Molecule type: mRNA
A;Residues: 1-177 <CAR>
A;Cross-references: GB:X53296; NID:g32578; PIDN:CAA37386.1; PID:g32579
A;Note: parts of this sequence, including the amino end of the mature protein, were conf
R;Eisenberg, S.P.; Evans, R.J.; Arend, W.P.; Verderber, E.; Brewer, M.T.; Hannum, C.H.;
Nature 343, 341-346, 1990
A;Title: Primary structure and functional expression from complementary DNA of a human i
A;Reference number: S08160; MUID:90136921
A;Accession: S08160
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-177 <EI2>
A;Cross-references: GB:X52015; NID:g32576; PIDN:CAA36262.1; PID:g32577
R;Hannum, C.H.; Wilcox, C.J.; Arend, W.P.; Joslin, F.G.; Dripps, D.J.; Heimdal, P.L.; Ar
Nature 343, 336-340, 1990
A;Title: Interleukin-1 receptor antagonist activity of a human interleukin-1 inhibitor.
A;Reference number: S08159; MUID:90136920
A;Accession: S08159
A;Molecule type: protein
A;Residues: 26-75;97-108;110-116;120-131;163-176 <HAN>
R;Bienkowski, M.J.; Eessalu, T.E.; Berger, A.E.; Truesdell, S.E.; Shelly, J.A.; Laborde,
J. Biol. Chem. 265, 14505-14511, 1990
A;Title: Purification and characterization of interleukin 1 receptor level antagonist pr
A;Reference number: A37822; MUID:90354444
A;Accession: A37822
A;Molecule type: protein
A;Residues: 26-52;70-77;122-127;170-175 <BIE>
A;Experimental source: culture medium, PMA-stimulated THP-1 cells
C;Comment: For an alternative splice form, see PIR:A39386
C;Genetics:
A;Gene: GDB:IL1RN
A;Cross-references: GDB:125897; OMIM:147679
A;Map position: 2q14.2-2q14.2
A;Introns: 39/2; 69/1; 106/3
C;Superfamily: interleukin-1
C;Keywords: alternative splicing; cytokine receptor; extracellular protein; glycoprotein
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-177/Product: interleukin-1 receptor antagonist #status experimental <MAT>
F;109/Binding site: carbohydrate (Asn) (covalent) #status experimental


  Query Match             77.8%;  Score 35;  DB 2;  Length 177;
  Best Local Similarity   66.7%;  Pred. No. 5.9;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
             || | |:||
Db      59 YLQGPNVNL 67
```

```
RESULT   5
A39386
interleukin-1 receptor antagonist, long intracellular splice form - human
N;Contains: interleukin-1 receptor antagonist, short intracellular splice form
C;Species: Homo sapiens (man)
C;Date: 28-Feb-1992 #sequence_revision 11-Apr-1997 #text_change 26-May-2000
C;Accession: I37893; A39386
R;Muzio, M.; Polentarutti, N.; Sironi, M.; Poli, G.; De Gioia, L.; Introna, M.; Manto
J. Exp. Med. 182, 623-628, 1995
A;Title: Cloning and characterization of a new isoform of the interleukin 1 receptor
A;Reference number: I37893; MUID:95355865
A;Accession: I37893
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-180 <RES>
A;Cross-references: EMBL:X84348; NID:g1008970; PIDN:CAA59087.1; PID:g1008971
R;Haskill, S.; Martin, G.; Van Le, L.; Morris, J.; Peace, A.; Bigler, C.F.; Jaffe, G.
Proc. Natl. Acad. Sci. U.S.A. 88, 3681-3685, 1991
A;Title: cDNA cloning of an intracellular form of the human interleukin 1 receptor an
A;Reference number: A39386; MUID:91219436
A;Accession: A39386
A;Molecule type: mRNA
A;Residues: 1-3,25-180 <HAS>
A;Cross-references: GB:M55646; NID:g186291; PIDN:AAA59138.1; PID:g186292
C;Comment: For an alternative splice form, see PIR:A30368
C;Genetics:
A;Gene: GDB:IL1RN
A;Cross-references: GDB:125897; OMIM:147679
A;Map position: 2q14.2-2q14.2
C;Superfamily: interleukin-1
C;Keywords: alternative splicing; cytokine receptor
F;1-180/Product: interleukin-1 receptor antagonist, long intracellular splice form #s
F;1-3,25-180/Product: interleukin-1 receptor antagonist, short intracellular splice f


  Query Match             77.8%;  Score 35;  DB 2;  Length 180;
  Best Local Similarity   66.7%;  Pred. No. 6;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
             || | |:||
Db      62 YLQGPNVNL 70


RESULT   6
I64001
hypothetical protein HI0117 - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 31-Oct-1997
C;Accession: I64001
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630
A;Accession: I64001
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-83 <TIGR>
A;Cross-references: GB:U32697; GB:L42023; NID:g1573067; PID:g1573069; TIGR:HI0117


  Query Match             75.6%;  Score 34;  DB 2;  Length 83;
  Best Local Similarity   100.0%;  Pred. No. 4.2;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGANIN 8
             |||||||
Db       9 LSGANIN 15
```

```
RESULT   7
E71215
hypothetical protein PH1989 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C;Accession: E71215
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137
A;Accession: E71215
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-335 <KAW>
A;Cross-references: GB:AP000007; NID:g3236134; PID:g3258433
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1989
C;Superfamily: conserved hypothetical protein MTH887


  Query Match             75.6%;  Score 34;  DB 2;  Length 335;
  Best Local Similarity   66.7%;  Pred. No. 19;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANINL 9
          :| |||||:
Db     59 FLKGANINV 67


RESULT   8
E75352
glycine cleavage system P protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: E75352
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: E75352
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-949 <WHI>
A;Cross-references: GB:AE002021; GB:AE000513; NID:g6459573; PIDN:AAF11360.1; PID:g645958
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR1809
A;Map position: 1


  Query Match             75.6%;  Score 34;  DB 2;  Length 949;
  Best Local Similarity   75.0%;  Pred. No. 58;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANIN 8
          || |||:|
Db    668 YLDGANMN 675

RESULT   9
S36834
glycine dehydrogenase (decarboxylating) (EC 1.4.4.2) - Escherichia coli
C;Species: Escherichia coli
C;Date: 22-Jan-1994 #sequence_revision 10-Nov-1995 #text_change 08-Oct-1999
C;Accession: S36834; I41232; G65074
R;Okamura-Ikeda, K.; Ohmura, Y.; Fujiwara, K.; Motokawa, Y.
Eur. J. Biochem. 216, 539-548, 1993
```

```
A;Title: Cloning and nucleotide sequence of the gcv operon encoding the Escherichia c
A;Reference number: S36832; MUID:93387305
A;Accession: S36834
A;Molecule type: DNA
A;Residues: 1-957 <OKA>
A;Cross-references: EMBL:X73958; NID:g403342; PIDN:CAA52146.1; PID:g403345
R;Stauffer, L.T.; Fogarty, S.J.; Stauffer, G.V.
Gene 142, 17-22, 1994
A;Title: Characterization of the Escherichia coli gcv operon.
A;Reference number: I41231; MUID:94237484
A;Accession: I41232
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-957 <RES>
A;Cross-references: GB:L20872; NID:g304890; PIDN:AAA23867.1; PID:g304892
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: G65074
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-957 <BLAT>
A;Cross-references: GB:AE000373; GB:U00096; NID:g2367173; PIDN:AAC75941.1; PID:g17892
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: gcvP; gcvHP
C;Keywords: oxidoreductase; phosphoprotein; pyridoxal phosphate
F;708/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted


  Query Match             75.6%;  Score 34;  DB 2;  Length 957;
  Best Local Similarity   75.0%;  Pred. No. 59;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANIN 8
          || |||:|
Db    679 YLDGANMN 686


RESULT  10
T16734
hypothetical protein R12C12.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T16734
R;Favello, T.
submitted to the EMBL Data Library, July 1995
A;Description: The sequence of C. elegans cosmid R12C12.
A;Reference number: Z18568
A;Accession: T16734
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-979 <FAV>
A;Cross-references: EMBL:U23510; NID:g746453; PID:g746454; PIDN:AAC46780.1; CESP:R12C
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:R12C12.1
A;Introns: 52/3; 149/3; 325/2; 393/1; 470/3; 546/3; 619/3; 749/2; 827/1; 883/1


  Query Match             75.6%;  Score 34;  DB 2;  Length 979;
  Best Local Similarity   75.0%;  Pred. No. 60;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANIN 8
          || |||:|
Db    702 YLDGANMN 709


RESULT  11
A39521
```

```
glycine dehydrogenase (decarboxylating) (EC 1.4.4.2) precursor - chicken
N;Alternate names: glycine decarboxylase; P-protein
C;Species: Gallus gallus (chicken)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 05-Dec-1998
C;Accession: A39521; C39521; A27483
R;Kume, A.; Koyata, H.; Sakakibara, T.; Ishiguro, Y.; Kure, S.; Hiraga, K.
J. Biol. Chem. 266, 3323-3329, 1991
A;Title: The glycine cleavage system. Molecular cloning of the chicken and human glycine
A;Reference number: A39521; MUID:91131643
A;Accession: A39521
A;Molecule type: mRNA
A;Residues: 1-1003 <KUM>
A;Cross-references: GB:D90266
A;Accession: C39521
A;Molecule type: DNA
A;Residues: 1-13 <KU2>
A;Cross-references: GB:D90240
A;Note: part of this sequence was confirmed by protein sequencing
R;Fujiwara, K.; Okamura-Ikeda, K.; Motokawa, Y.
Biochem. Biophys. Res. Commun. 149, 621-627, 1987
A;Title: Amino acid sequence of the phosphopyridoxyl peptide from P-protein of the chick
A;Reference number: A27483; MUID:88106483
A;Accession: A27483
A;Molecule type: protein
A;Residues: 703-756 <FUJ>
C;Keywords: mitochondrion; oxidoreductase; phosphoprotein; pyridoxal phosphate
F;737/Binding site: pyridoxal phosphate (Lys) (covalent) #status experimental


  Query Match             75.6%;  Score 34;  DB 2;  Length 1003;
  Best Local Similarity   75.0%;  Pred. No. 62;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANIN 8
            || |||:|
Db    708 YLDGANMN 715


RESULT 12
JN0124
glycine dehydrogenase (decarboxylating) (EC 1.4.4.2) - human
N;Alternate names: glycine decarboxylase; P-protein
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-Nov-1999
C;Accession: JN0124
R;Kure, S.; Narisawa, K.; Tada, K.
Biochem. Biophys. Res. Commun. 174, 1176-1182, 1991
A;Title: Structural and expression analyses of normal and mutant mRNA encoding glycine d
A;Reference number: JN0124; MUID:91144593
A;Accession: JN0124
A;Molecule type: mRNA
A;Residues: 1-1020 <KUR>
A;Cross-references: GB:M63635; NID:g190286; PIDN:AAA36478.1; PID:g190287
A;Note: deletion of 756-Phe causes nonketotic hyperglycinemia
C;Genetics:
A;Gene: GDB:GLDC
A;Cross-references: GDB:128611; OMIM:238300
A;Map position: 9p22-9p22
C;Keywords: mitochondrion; oxidoreductase; phosphoprotein; pyridoxal phosphate
F;761-768/Region: glycine-rich
F;754/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted


  Query Match             75.6%;  Score 34;  DB 2;  Length 1020;
  Best Local Similarity   75.0%;  Pred. No. 63;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANIN 8
            || |||:|
Db    725 YLDGANMN 732


RESULT 13
```

```
B39521
glycine dehydrogenase (decarboxylating) (EC 1.4.4.2) precursor - human
N;Alternate names: glycine decarboxylase; P-protein
C;Species: Homo sapiens (man)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 05-Nov-1999
C;Accession: B39521
R;Kume, A.; Koyata, H.; Sakakibara, T.; Ishiguro, Y.; Kure, S.; Hiraga, K.
J. Biol. Chem. 266, 3323-3329, 1991
A;Title: The glycine cleavage system. Molecular cloning of the chicken and human glyc
A;Reference number: A39521; MUID:91131643
A;Accession: B39521
A;Molecule type: mRNA
A;Residues: 1-1020 <KUM>
A;Cross-references: GB:M64590; GB:J05742; NID:g190208; PIDN:AAA36463.1; PID:g190209
C;Genetics:
A;Gene: GDB:GLDC
A;Cross-references: GDB:128611; OMIM:238300
A;Map position: 9p22-9p22
C;Keywords: mitochondrion; oxidoreductase; phosphoprotein; pyridoxal phosphate
F;754/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted


  Query Match             75.6%;  Score 34;  DB 2;  Length 1020;
  Best Local Similarity   75.0%;  Pred. No. 63;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANIN 8
            || |||:|
Db    725 YLDGANMN 732


RESULT 14
S50917
aminomethyltransferase (EC 2.1.2.10) - yeast (Saccharomyces cerevisiae)
N;Alternate names: glycin cleavage system protein P; protein YM9646.01; protein YMR18
C;Species: Saccharomyces cerevisiae
C;Date: 10-Feb-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C;Accession: S50917; S59810; S70896
R;Pearson, D.; Bowman, S.
submitted to the EMBL Data Library, January 1995
A;Reference number: S50917
A;Accession: S50917
A;Molecule type: DNA
A;Residues: 1-1034 <PEA>
A;Cross-references: EMBL:Z47815; NID:g642280; PIDN:CAA87810.1; PID:g642281; MIPS:YMR1
R;Sinclair, D.A.; Dawes, I.W.
submitted to the EMBL Data Library, February 1995
A;Description: Cloning and expression of glycine decarboxylase from Saccharomyces cer
A;Reference number: S59809
A;Accession: S59810
A;Molecule type: DNA
A;Residues: 1-1034 <SIN>
A;Cross-references: EMBL:U20641; NID:g676869; PIDN:AAB18933.1; PID:g676871
R;Sinclair, D.A.; Hong, S.P.; Dawes, I.W.
Mol. Microbiol. 19, 611-623, 1996
A;Title: Specific induction by glycine of the gene for the P-subunit of glycine decar
A;Reference number: S70896; MUID:96228709
A;Accession: S70896
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-77,'V',79-121,'VS',124-247,'L',249-381,'T',383-517,'H',519-1034 <SIW>
A;Cross-references: EMBL:U20641
C;Genetics:
A;Gene: GCV2; GSD2
A;Map position: 13R
C;Keywords: phosphoprotein; pyridoxal phosphate; transferase
F;773/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted


  Query Match             75.6%;  Score 34;  DB 2;  Length 1034;
  Best Local Similarity   75.0%;  Pred. No. 64;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy       1 YLSGANIN 8
             || |||:|
Db     744 YLDGANMN 751


RESULT  15
E71551
probable monooxygenase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C;Accession: E71551
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac
A;Reference number: A71570; MUID:99000809
A;Accession: E71551
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-507 <ARN>
A;Cross-references: GB:AE001288; GB:AE001273; NID:g3328545; PIDN:AAC67739.1; PID:g332854
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: mhpA


  Query Match             73.3%;  Score 33;  DB 2;  Length 507;
  Best Local Similarity   66.7%;  Pred. No. 48;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
             |||| | |:
Db     290 YLSGVNSNI 298


Search completed: December 16, 2000, 01:51:16
Job time: 7637 sec
```

```
                       GenCore version 4.5
              Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 00:51:14 ; Search time 108.84 Seconds
                                              (without alignments)
                                              2.827 Million cell updates/sec

Title:           US-09-529-121-4
Perfect score:   45
Sequence:        1 YLSGANINL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters:      268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_36:*
               1:  /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
               2:  /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
               3:  /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
               4:  /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
               5:  /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
               6:  /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
               7:  /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
               8:  /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
               9:  /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
              10:  /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
              11:  /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
              12:  /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
              13:  /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
              14:  /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
              15:  /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
              16:  /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
              17:  /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
              18:  /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
              19:  /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
              20:  /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
              21:  /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.

                             SUMMARIES
                %
Result        Query
  No.   Score Match Length DB  ID                    Description
 -----------------------------------------------------------------------
    1     45  100.0      9 20  Y09528                Carcinoembryonic a
    2     43   95.6      9 18  W39723                Human carcina-embr
    3     43   95.6      9 19  W77134                CEA synthetic pept
    4     43   95.6      9 19  W70045                CEA derived HLA-A2
    5     43   95.6      9 20  Y47655                Immunogenic peptid
    6     43   95.6      9 20  Y09525                Carcinoembryonic a
    7     43   95.6     10 20  Y46555                Immunogenic peptid
    8     43   95.6    107 20  W86133                Protein sequence o
    9     43   95.6    178 10  P93499                Sequence of carcin
   10     43   95.6    468 16  R77436                BGP (1-314)/CEA (4
   11     43   95.6    493 16  R77435                BGP (1-314)/CEA (4
   12     43   95.6    509 16  R77437                BGP (1-314)/CEA (4
```

```
   13     43   95.6    511 16  R77438                BGP (1-314)/CEA (4
   14     43   95.6    642 15  R60619                Carcinoembryonic a
   15     43   95.6    663 17  R98519                Immunogenic carcin
   16     43   95.6    698 9  P81229                Carcinoembryonic a
   17     43   95.6    698 16  R65168                Carcinoembryonic a
   18     43   95.6    698 18  W22844                Human carcinoembry
   19     43   95.6    702 9  P81222                Carcinoembryonic a
   20     43   95.6    702 10  P94014                Carcinoembryonic c
   21     43   95.6    702 10  P93999                Amino acid sequenc
   22     43   95.6    702 15  R54713                Carcinoembryonic a
   23     43   95.6    702 17  W06872                Carcinoembryonic a
   24     43   95.6    702 20  W83137                CEA protein.  Homo
   25     43   95.6    734 17  W00182                Carcinoembryonic a
   26     40   88.9      9 20  Y09527                Carcinoembryonic a
   27     40   88.9      9 21  Y54173                HLA binding peptid
   28     38   84.4      9 20  Y09526                Carcinoembryonic a
   29     36   80.0      9 17  W00680                Peptide comprising
   30     35   77.8     90 10  P93625                Sequence of interl
   31     35   77.8     90 14  R35484                IL-li fragment.  H
   32     35   77.8    152 13  R27495                Native IL-lra poly
   33     35   77.8    152 14  R35486                IL-1 inhibitor (IL
   34     35   77.8    153 18  W22894                Recombinant human
   35     35   77.8    153 19  W61149                Recombinant human
   36     35   77.8    159 12  R15262                Variant Il-1 cytok
   37     35   77.8    159 16  R73642                iCIL-1ra.  Homo sa
   38     35   77.8    159 19  W80777                Human intracellula
   39     35   77.8    159 19  W37787                Recombinant interl
   40     35   77.8    159 20  Y43532                Cytoplasmic form o
   41     35   77.8    159 20  Y33275                icIRAP conserved p
   42     35   77.8    159 20  Y33277                IcIRAP conserved p
   43     35   77.8    159 20  Y28287                Polypeptide sequen
   44     35   77.8    159 20  W73946                IL-1 receptor anta
   45     35   77.8    159 20  W83008                Intracellular inte

                              ALIGNMENTS

RESULT   1
Y09528
ID   Y09528 standard; peptide; 9 AA.
XX
AC   Y09528;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:4.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
```

```
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;


  Query Match             100.0%;  Score 45;  DB 20;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
           |||||||||
Db       1 ylsganinl 9


RESULT   2
W39723
ID   W39723 standard; peptide; 9 AA.
XX
AC   W39723;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Human carcina-embryonic antigen (CEA) peptide (pos. 571-579).
XX
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
XX
OS   Homo sapiens.
XX
PN   WO9741440-A1.
XX
PD   06-NOV-1997.
XX
PF   28-APR-1997;   97WO-NL00229.
XX
PR   23-DEC-1996;   96EP-0203670.
PR   26-APR-1996;   96EP-0201145.
XX
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
PI   Kast WM,  Melief CJM,  Offringa R,  Toes REM,  Van Der Burg SH;
XX
DR   WPI; 1997-549891/50.
XX
PT   Method of selecting T cell peptide epitope(s) - by measuring the
PT   stability of HLA class I-peptide complexes on intact B cells
XX
PS   Example 3; Page 85; 109pp; English.
XX
CC   Peptides W39430-W39734 are used in a novel method for the selection of
CC   immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC   method involves the identification of peptide sequences capable of
CC   binding to an HLA (human leukocyte antigen) class I molecule and
CC   measuring the binding of this epitope peptide to the HLA class I peptide.
CC   The stability of binding of the peptide and MHC (major histocompatablilty
CC   complex) class I molecule is measured on intact human B cells carrying
CC   the MHC molecule at their cell surfaces. The method can be used to select
CC   peptide epitopes for generating vaccines against a disease associated
```

```
CC   with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC   especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC   immune responses. Peptide W39723 is derived from the human
CC   carcino-embryonic antigen (CEA) and has the ability to bind to the human
CC   MHC Class I allele HLA-A2.1.
XX
SQ   Sequence   9 AA;


  Query Match             95.6%;  Score 43;  DB 18;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
           ||||||:||
Db       1 ylsganlnl 9


RESULT   3
W77134
ID   W77134 standard; peptide; 9 AA.
XX
AC   W77134;
XX
DT   16-NOV-1998  (first entry)
XX
DE   CEA synthetic peptide epitope 1.
XX
KW   Tyrosinase; tyrosinase cytotoxic lymphocyte response;
KW   cytotoxic T lymphocyte; cysteine-depleted; melanoma.
XX
OS   Synthetic.
XX
PN   WO9833810-A2.
XX
PD   06-AUG-1998.
XX
PF   29-JAN-1998;   98WO-US01592.
XX
PR   30-JAN-1997;   97US-0037781.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI   Engelhard VH,  Hunt DF,  Kittlesen D,  Slingluff CL;
XX
DR   WPI; 1998-437388/37.
XX
PT   Disease specific immunogen  - comprises disease specific cytotoxic T
PT   lymphocyte epitope used to elicit melanoma specific CTL response
XX
PS   Disclosure; Page 27; 93pp; English.
XX
CC   The peptide epitope W77119-W77138 were created for human tumour-specific
CC   cytotoxic T lymphocyte response.  These peptides are are cysteine-
CC   depleted mutants of a native disease-specific CTL epitope.  The cysteine-
CC   depleted CTL epitopes elicit a stronger or more specific CTL response
CC   than the native epitope.  The epitopes can be used in a disease-specific
CC   immunogen to protect a mammal against disease in particular melanomas.
CC   The peptides may also be used to screen a sample for the presence of
CC   an antigen with the same epitope, or with a different cross-reactive
CC   epitope.
XX
SQ   Sequence   9 AA;


  Query Match             95.6%;  Score 43;  DB 19;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
           ||||||:||
Db       1 ylsganlnl 9
```

```
RESULT   4
W70045
ID   W70045 standard; peptide; 9 AA.
XX
AC   W70045;
XX
DT   22-OCT-1998  (first entry)
XX
DE   CEA derived HLA-A2.1 binding peptide 2 (residues 605-613).
XX
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW   antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW   viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW   fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9833888-A1.
XX
PD   06-AUG-1998.
XX
PF   30-JAN-1998;   98WO-US01959.
XX
PR   31-JAN-1997;   97US-0036696.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Celis E,  Sette A,  Sidney J,  Southwood S,  Tsai V;
XX
DR   WPI; 1998-437445/37.
XX
PT   Production of antigen-specific cytotoxic T cells - by incubating
PT   immunogenic peptide(s) from antigen that binds class I major
PT   histocompatibility complex molecules with pre-treated antigen
PT   presenting cells
XX
PS   Example 6; Page 75; 104pp; English.
XX
CC   Sequences shown in W70044 to W70052 represent peptides derived from
CC   carcinoembryonic antigen (CEA). The peptides can bind to a human
CC   leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method
CC   of invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC   vitro. The method comprises contacting immunogenic peptides from an
CC   antigen that binds class I major histocompatibility complex (MHC)
CC   molecules with antigen presenting cells (APCs) pretreated with
CC   pretreatment growth factors, and incubating the APCs with purified CD8
CC   cells in the presence of at least 2 incubation growth factors, thereby
CC   producing antigen-specific CTLs. A method for specifically killing
CC   target cells in a human patient is also provided which comprises
CC   obtaining a fluid sample containing CTLs from a patient, contacting the
CC   cytotoxic T cells with APCs pretreated with pre-treatment growth factors,
CC   where the APCs comprise class I MHC molecules. The pretreated APCs are
CC   incubated with the cytotoxic growth factors, thereby producing activated
CC   CTLs which are contacted with a carrier to form a composition. The
CC   composition can then be administered to the patient. The activated CTLs
CC   can be used for treating cancers, immune disorders, viral infections,
CC   AIDS, hepatitis, bacterial infection, fungal infection, malaria or
CC   tuberculosis.
XX
SQ   Sequence   9 AA;

  Query Match             95.6%;  Score 43;  DB 19;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
           ||||||:||
Db       1 ylsganlnl 9
```

```
RESULT   5
Y47655
ID   Y47655 standard; Peptide; 9 AA.
XX
AC   Y47655;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2266.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 118; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;

  Query Match             95.6%;  Score 43;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
           ||||||:||
Db       1 ylsganlnl 9
```

```
RESULT   6
Y09525
ID   Y09525 standard; peptide; 9 AA.
XX
AC   Y09525;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist CAP-1.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 1; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             95.6%;  Score 43;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              ||||||:||
Db          1 ylsganlnl 9


RESULT   7
Y46555
ID   Y46555 standard; Peptide; 10 AA.
XX
AC   Y46555;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #1166.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
```

```
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 76; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   10 AA;

  Query Match             95.6%;  Score 43;  DB 20;  Length 10;
  Best Local Similarity   88.9%;  Pred. No. 0.0079;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGANINL 9
              ||||||:||
Db          2 ylsganlnl 10


RESULT   8
W86133
ID   W86133 standard; Protein; 107 AA.
XX
AC   W86133;
XX
DT   03-MAR-1999  (first entry)
XX
DE   Protein sequence of vaccine 2 708 Vl.
XX
KW   Non-immunogenic; epitope; T-cell; immunogenicity; immune system; SK;
KW   immunoglobulin; therapeutic; streptokinase; vaccine; 708.
```

```
XX
OS   Homo sapiens.
XX
PN   WO9852976-A1.
XX
PD   26-NOV-1998.
XX
PF   21-MAY-1998;   98WO-GB01473.
XX
PR   14-APR-1998;   98GB-0007751.
PR   21-MAY-1997;   97GB-0010480.
PR   31-JUL-1997;   97GB-0016197.
PR   28-NOV-1997;   97GB-0025270.
PR   02-DEC-1997;   97US-0067235.
XX
PA   (BIOV-) BIOVATION LTD.
XX
PI   Carr FJ;
XX
DR   WPI; 1999-045301/04.
XX
PT   Reducing immunogenicity of proteins - by modifying the amino acid
PT   sequence of the protein to eliminate potential epitopes for T-cells
PT   of a given species
XX
PS   Example 4; Fig 19; 77pp; English.
XX
CC   The invention relates to a method for the production of non-immunogenic
CC   proteins. The method comprises determining at least part of the amino
CC   acid sequence of the protein; (b) identifying in the amino acid sequence
CC   one or more potential epitopes for T-cells (T-cell epitopes) of the given
CC   species; and (c) modifying the amino acid sequence to eliminate at least
CC   one of the T-cell epitopes identified in step (b) thereby to eliminate or
CC   reduce the immunogenicity of the protein when exposed to the immune
CC   system of the given species. A method of analysing a pre-existing protein
CC   to predict the basis for immunogenic responses is also provided. The
CC   methods can be used particularly for reducing the immunogenicity of
CC   immunoglobulins or therapeutic proteins, e.g. Streptokinase (SK). The
CC   products can be used for diagnosis and therapy. The present sequence
CC   represents the protein sequence of vaccine 2 708 V1.
XX
SQ   Sequence   107 AA;

  Query Match             95.6%;  Score 43;  DB 20;  Length 107;
  Best Local Similarity   88.9%;  Pred. No. 0.11;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YLSGANINL 9
            ||||||:||
Db     89 ylsganlnl 97


RESULT  9
P93499
ID   P93499 standard; protein; 178 AA.
XX
AC   P93499;
XX
DT   08-MAY-1990  (first entry)
XX
DE   Sequence of carcinoembryonic antigen domain III.
XX
KW   Carcinoembryonic antigen; domain III; domain A; domain B.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..89
FT                   /note="domain A"
FT   Domain          90..178
FT                   /note="domain B"
XX
PN   EP343946-A.
XX
```

```
PD   29-NOV-1989.
XX
PF   24-MAY-1989;   89EP-0305232.
XX
PR   25-MAY-1988;   88US-0198289.
XX
PA   (CITY ) CITY OF HOPE.
XX
PI   Shively JE;
XX
DR   WPI; 1989-349991/48.
DR   N-PSDB; N92449.
XX
PT   Carcinoembryonic antigen fragments - used in assays to determine the
PT   presence and amt. of the antigen in samples also contg. related antigens.
XX
PS   Disclosure; page 4; 15pp; English.
XX
CC   CEA fragments can be used in assays to determine the presence and amt. of
CC   CEA in samples which also may contain related antigens including its
CC   normal cross-reacting antigen or the 128 kD antigen.
XX
SQ   Sequence   178 AA;

  Query Match             95.6%;  Score 43;  DB 10;  Length 178;
  Best Local Similarity   88.9%;  Pred. No. 0.2;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YLSGANINL 9
            ||||||:||
Db    107 ylsganlnl 115


RESULT  10
R77436
ID   R77436 standard; Protein; 468 AA.
XX
AC   R77436;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-643) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..468
FT                   /note= "CEA (490-643)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
```

```
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 16; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   468 AA;


  Query Match             95.6%;  Score 43;  DB 16;  Length 468;
  Best Local Similarity   88.9%;  Pred. No. 0.58;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YLSGANINL 9
            ||||||:||
Db    396 ylsganlnl 404


RESULT  11
R77435
ID   R77435 standard; Protein; 493 AA.
XX
AC   R77435;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-C-terminal) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..493
FT                   /note= "CEA (490-C-terminal)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
```

```
XX
PS   Claim 15; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   493 AA;


  Query Match             95.6%;  Score 43;  DB 16;  Length 493;
  Best Local Similarity   88.9%;  Pred. No. 0.62;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YLSGANINL 9
            ||||||:||
Db    396 ylsganlnl 404


RESULT  12
R77437
ID   R77437 standard; Protein; 509 AA.
XX
AC   R77437;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-644)/BGP (391-430) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..469
FT                   /note= "CEA (490-644)"
FT   Protein         470..509
FT                   /note= "BGP (391-430)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
```

```
XX
PS   Claim 17; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   509 AA;


  Query Match             95.6%;  Score 43;  DB 16;  Length 509;
  Best Local Similarity   88.9%;  Pred. No. 0.64;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
           ||||||:||
Db     396 ylsganlnl 404


RESULT  13
R77438
ID   R77438 standard; Protein; 511 AA.
XX
AC   R77438;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-642)/BGP (387-430) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..467
FT                   /note= "CEA (490-642)"
FT   Protein         468..511
FT                   /note= "BGP (387-430)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
```

```
XX
PS   Claim 18; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   511 AA;


  Query Match             95.6%;  Score 43;  DB 16;  Length 511;
  Best Local Similarity   88.9%;  Pred. No. 0.64;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
           ||||||:||
Db     396 ylsganlnl 404


RESULT  14
R60619
ID   R60619 standard; Protein; 642 AA.
XX
AC   R60619;
XX
DT   10-MAY-1995  (first entry)
XX
DE   Carcinoembryonic antigen glycoprotein.
XX
KW   Carcinoembryonic antigen; CEA; neoplastic diseases.
XX
OS   Homo sapiens.
XX
PN   EP618292-A.
XX
PD   05-OCT-1994.
XX
PF   15-MAR-1994;   94EP-0103986.
XX
PR   25-MAR-1993;   93EP-0810214.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Mach J,  Pelegrin A,  Terskikh A;
XX
DR   WPI; 1994-304461/38.
DR   N-PSDB; Q71567.
XX
PT   Carcinoembryonic antigen (CEA) derivs - useful as reagents in
PT   immunoassay for diagnosis of neoplastic diseases
XX
PS   Claim 2; Page 15; 30pp; English.
XX
CC   Q71567 is the DNA sequence which encodes carcinoembryonic antigen (CEA)
CC   R60619. CEA is free from cross-reactive CEA-like antigens, it is
CC   antigenically indistinguishable from the solution form of CEA shed from
CC   tumour cells, and it is devoid of ethanolamine. R60619 can be used in a
CC   reagent composition for detecting neoplastic diseases in biological
CC   samples, or in an immunoassay process where it can specifically detect
CC   the presence of tumour cells in a biological sample e.g. blood.
XX
SQ   Sequence   642 AA;
```

```
  Query Match             95.6%;  Score 43;  DB 15;  Length 642;
  Best Local Similarity   88.9%;  Pred. No. 0.83;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
           ||||||:||
Db     571 ylsganlnl 579


RESULT  15
R98519
ID   R98519 standard; Protein; 663 AA.
XX
AC   R98519;
XX
DT   13-NOV-1996  (first entry)
XX
DE   Immunogenic carcinoembryonic antigen.
XX
KW   Carcinoembryonic antigen; immunogen; breast cancer; lung cancer;
KW   colon cancer; therapy; immunotherapy; vaccine; baculovirus; vector;
KW   Spodoptera frugiperda; insect; pA9080 AcNPV-CEA.
XX
OS   Chimeric Autographa californica nuclear polyhedrosis virus;
OS   Chimeric Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..18
FT                   /label= Sig_peptide
FT                   /note= "AcNPV 61k protein signal peptide"
FT   Protein         18..663
FT                   /label= Mat_protein
FT                   /note= "amino acids 1-3 of the mature protein
FT                          are derived from the baculovirus vector"
XX
PN   WO9532286-A2.
XX
PD   30-NOV-1995.
XX
PF   19-MAY-1995;   95WO-US06373.
XX
PR   20-MAY-1994;   94US-0246981.
XX
PA   (MICR-) MICROGENESYS INC.
XX
PI   Hackett C,  Smith G,  Volvovitz F;
XX
DR   WPI; 1996-020581/02.
DR   N-PSDB; T36495.
XX
PT   Immunogenic carcinoembryonic antigen produced using insect cell
PT   baculovirus expression system - useful in cancer therapy
XX
PS   Claim 9; Page 50-53; 61pp; English.
XX
CC   A recombinant, soluble, immunogenic carcinoembryonic antigen
CC   (rCEA) (R98519) is encoded by vector pA9080 AcNPV-CEA (see also
CC   T36495) in which a modified human CEA gene is joined to a
CC   baculovirus signal sequence under control of a polyhedrin
CC   promoter.  The baculovirus signal peptide directs translation of
CC   rCEA into the insect cell glycosylation pathway.  rCEA can be
CC   produced at high levels in Sf900+ insect cells grown in serum-free
CC   media, and isolated to a purity of over 95%.  It is used as an
CC   immunogen in humans to protect against cancer, partic. breast,
CC   lung or colon cancer.
XX
SQ   Sequence   663 AA;

  Query Match             95.6%;  Score 43;  DB 17;  Length 663;
  Best Local Similarity   88.9%;  Pred. No. 0.86;
```

```
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
           ||||||:||
Db     592 ylsganlnl 600


Search completed: December 16, 2000, 00:51:14
Job time: 18774 sec
```

```
                           GenCore version 4.5
                 Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 04:22:13 ; Search time 111.26 Seconds
                                              (without alignments)
                                              7.553 Million cell updates/sec

Title:           US-09-529-121-4
Perfect score:   45
Sequence:        1 YLSGANINL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters:      436

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       SPTREMBL_14:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                  %
Result           Query
  No.   Score   Match Length DB  ID                     Description
  -------------------------------------------------------------------------
    1      15    33.3      8  3  P87225                 P87225 saccharomyc
    2      15    33.3      8  4  Q9UL56                 Q9ul56 homo sapien
    3      15    33.3      8 13  P82079                  P82079 limnodynast
    4      14    31.1      7 12  Q86871                  Q86871 cauliflower
    5      14    31.1      8  2  Q9X3K1                 Q9x3k1 prochloroco
    6      14    31.1      8 12  Q66807                  Q66807 echovirus 2
    7      14    31.1      8 13  Q9PS69                  Q9ps69 gallus gall
    8      14    31.1      9  2  Q9R7E8                 Q9r7e8 escherichia
    9      14    31.1      9  6  Q9TRS0                 Q9trs0 oryctolagus
   10      14    31.1      9  7  Q31415                 Q31415 gallus gall
   11      14    31.1      9 11  O35953                  O35953 mus musculu
   12      14    31.1      9 12  Q89491                  Q89491 murine minu
   13      13    28.9      7 12  Q07624                  Q07624 rous sarcom
   14      13    28.9      8 13  Q90498                  Q90498 erythrura g
   15      13    28.9      8 13  Q91098                  Q91098 manorina me
   16      13    28.9      9  2  Q48686                 Q48686 lactococcus
   17      13    28.9      9  2  Q44001                 Q44001 aeromonas e
   18      13    28.9      9  2  Q44377                 Q44377 aeromonas t
   19      13    28.9      9  2  Q44468                 Q44468 aeromonas v
   20      13    28.9      9  2  Q43928                 Q43928 aeromonas c
   21      13    28.9      9  8  O78337                 O78337 caloglossa
   22      13    28.9      9  8  Q9TLD0                 Q9tld0 bostrychia
   23      13    28.9      9  8  Q9TJ87                 Q9tj87 caloglossa
   24      13    28.9      9  8  Q9TJ85                 Q9tj85 caloglossa
   25      13    28.9      9  8  Q9T389                 Q9t389 caloglossa
   26      13    28.9      9  8  Q9T388                 Q9t388 caloglossa
   27      13    28.9      9  8  Q9T387                 Q9t387 bostrychia
   28      13    28.9      9 10  O81962                  O81962 caloglossa
   29      13    28.9      9 10  O81964                  O81964 caloglossa
   30      13    28.9      9 10  O81966                  O81966 caloglossa
   31      13    28.9      9 10  O81968                  O81968 caloglossa
   32      13    28.9      9 10  O82778                  O82778 caloglossa
   33      12    26.7      8  4  Q15901                 Q15901 homo sapien
   34      12    26.7      8  4  Q9UMH9                 Q9umh9 homo sapien
   35      12    26.7      8  5  O02032                 O02032 lytechinus
   36      12    26.7      8  7  Q95213                 Q95213 oryctolagus
   37      12    26.7      9  2  Q9R635                 Q9r635 chlamydia t
   38      12    26.7      9  4  Q14715                 Q14715 homo sapien
   39      12    26.7      9  4  Q9UCN5                 Q9ucn5 homo sapien
   40      12    26.7      9  5  Q27396                 Q27396 babesia bov
   41      12    26.7      9  6  Q28112                 Q28112 bos taurus
   42      12    26.7      9  6  Q9TRW2                 Q9trw2 oryctolagus
   43      11    24.4      7  2  Q47505                 Q47505 escherichia
   44      11    24.4      7 12  Q9YQ10                  Q9yq10 porcine tra
   45      11    24.4      8  2  Q47273                 Q47273 escherichia
   46      11    24.4      8  2  P77556                 P77556 escherichia
   47      11    24.4      8  2  Q9RQ57                 Q9rq57 buchnera ap
   48      11    24.4      8  2  Q9RQ49                 Q9rq49 buchnera ap
   49      11    24.4      8  4  Q15889                 Q15889 homo sapien
   50      11    24.4      8  4  Q9UCN4                 Q9ucn4 homo sapien
   51      11    24.4      8  5  Q94695                 Q94695 physarum po
   52      11    24.4      8  5  Q9VVJ2                 Q9vvj2 drosophila
   53      11    24.4      8  6  Q9XSY1                 Q9xsy1 canis famil
   54      11    24.4      8 11  Q9QVF4                  Q9qvf4 rattus sp.
   55      11    24.4      8 12  Q83332                  Q83332 murine hepa
   56      11    24.4      8 12  Q85562                  Q85562 moloney mur
   57      11    24.4      8 13  P82082                  P82082 limnodynast
   58      11    24.4      8 13  P82083                  P82083 limnodynast
   59      11    24.4      9  1  Q50832                 Q50832 methanococc
   60      11    24.4      9  2  Q57328                 Q57328 aeromonas s
   61      11    24.4      9  2  O30790                 O30790 erwinia amy
   62      11    24.4      9  2  Q9R7H9                 Q9r7h9 haemophilus
   63      11    24.4      9  2  Q9R5M1                 Q9r5m1 staphylococ
   64      11    24.4      9  4  P78484                 P78484 homo sapien
   65      11    24.4      9  4  Q9UQA3                 Q9uqa3 homo sapien
   66      11    24.4      9  5  Q9TWD6                 Q9twd6 leptinotars
   67      11    24.4      9 13  Q92009                  Q92009 gallus gall
   68      10    22.2      5 10  Q99007                  Q99007 hordeum vul
   69      10    22.2      7 11  Q63480                  Q63480 rattus norv
   70      10    22.2      7 12  Q67113                  Q67113 influenza a
   71      10    22.2      8  2  P72221                 P72221 pseudomonas
   72      10    22.2      8  2  Q9R7T2                 Q9r7t2 escherichia
   73      10    22.2      8  2  Q9R4M3                 Q9r4m3 enterococcu
   74      10    22.2      8  4  Q16428                 Q16428 homo sapien
   75      10    22.2      8  4  Q9Y4J4                 Q9y4j4 homo sapien


                                  ALIGNMENTS


RESULT   1
P87225
ID    P87225      PRELIMINARY;      PRT;     8 AA.
AC    P87225;
DT    01-JUL-1997 (TrEMBLrel. 04, Created)
DT    01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT    01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE    GIN11 PROTEIN (FRAGMENT).
GN    GIN11.
OS    Saccharomyces cerevisiae (Baker's yeast).
OC    Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC    Saccharomycetaceae; Saccharomyces.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wedler H., Wedler E., Scharfe M., Wambutt R.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z73169; CAA97518.2; -.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  1019 MW;  4E21A9C449D5B73B CRC64;

  Query Match             33.3%;  Score 15;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLS 3
            |||
Db        1 YLS 3

RESULT   2
Q9UL56
ID   Q9UL56      PRELIMINARY;      PRT;`     8 AA.
AC   Q9UL56;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MUTANT NADH-CYTOCHROME B5 REDUCTASE (EC 1.6.2.2) (FRAGMENT).
GN   DIA1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fukumaki Y., Higasa K.;
RT   "Two novel mutations in Thai patients with hereditary
RT   methemoglobinemia types I and II: a subtle amino acid change causes
RT   instability of NADH-cytochrome b5 reductase.";
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF061830; AAF06818.1; -.
KW   Oxidoreductase.
FT   NON_TER       1      1
FT   VARIANT       9      9       -> R.
SQ   SEQUENCE   8 AA;  888 MW;  76C5B73B5051F6D8 CRC64;

  Query Match             33.3%;  Score 15;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLS 3
            |||
Db        5 YLS 7

RESULT   3
P82079
ID   P82079      PRELIMINARY;      PRT;     8 AA.
AC   P82079;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   DYNASTIN 1.
OS   Limnodynastes interioris (Giant banjo frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Myobatrachidae;
OC   Limnodynastes.
RN   [1]
RP   SEQUENCE, AND MASS SPECTROMETRY.
RC   TISSUE=TIBIAL GLAND;
RA   Raftery M.J., Bradford A.M., Bowie J.H., Wallace J.C., Tyler M.J.;
RT   "Peptides from Australian frogs. The structure of the dynastins from
RT   the banjo frogs Limnodynastes interioras, Limnodynastes dumerilii and
RT   Limnodynastes terraereginae.";
RL   Aust. J. Chem. 46:833-842(1993).
CC   -!- MASS SPECTROMETRY: MW=729; METHOD=FAB.
KW   Amphibian skin.
SQ   SEQUENCE   8 AA;  729 MW;  7C28772865B72728 CRC64;

  Query Match             33.3%;  Score 15;  DB 13;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        2 LSGANI 7
            |||  :
Db        3 LSGLGL 8

RESULT   4
Q86871
ID   Q86871      PRELIMINARY;      PRT;     7 AA.
AC   Q86871;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   VIRION STRUCTURAL PROTEIN (FRAGMENT).
GN   GENE III.
OS   Cauliflower mosaic virus.
OC   Viruses; Retroid viruses; Caulimovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=11/3-7;
RX   MEDLINE; 95053898.
RA   al-Kaff N., Covey S.N.;
RT   "Variation in biological properties of cauliflower mosaic virus
RT   clones.";
RL   J. Gen. Virol. 75:3137-3145(1994).
DR   EMBL; S75948; CAB33416.1; -.
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  744 MW;  672054444DC5B030 CRC64;

  Query Match             31.1%;  Score 14;  DB 12;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        2 LSGANI 7
            :|  ||
Db        1 ISANNI 6

RESULT   5
Q9X3K1
ID   Q9X3K1      PRELIMINARY;      PRT;     8 AA.
AC   Q9X3K1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CYTOCHROME B (FRAGMENT).
GN   PETB.
OS   Prochlorococcus sp.
OC   Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC   Prochlorococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Urbach E., Chisholm S.W.;
RT   "Genetic diversity in Prochlorococcus populations flow cytometrically
RT   sorted from the Sargasso Sea and Gulf Stream.";
RL   Limnol. Oceanog. 43:1615-1630(1998).
DR   EMBL; AF070193; AAD23233.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  799 MW;  10376865B72866D3 CRC64;
```

```
  Query Match             31.1%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSG 4
            |||
Db        4 LSG 6


RESULT   6
Q66807
ID   Q66807      PRELIMINARY;      PRT;     8 AA.
AC   Q66807;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   5'UTR IN IRES SEQUENCES (ISOLATE TH222) (FRAGMENT).
OS   Echovirus 25.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Enterovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TH222;
RA   Bailly J.L., Borman A.M., Peigue-Lafeuille H., Kean K.M.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X90724; CAA62259.1; -.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  821 MW;  EFC1B5A2D6DDD876 CRC64;


  Query Match             31.1%;  Score 14;  DB 12;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 GANIN 8
            || ::
Db        2 GAQVS 6


RESULT   7
Q9PS69
ID   Q9PS69      PRELIMINARY;      PRT;     8 AA.
AC   Q9PS69;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   LOW DENSITY LIPOPROTEIN RECEPTOR-RELATED PROTEIN.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92011685.
RA   Stifani S., Barber D.L., Aebersold R., Steyrer E., Shen X., Nimpf J.,
RA   Schneider W.J.;
RT   "The laying hen expresses two different low density lipoprotein
RT   receptor-related proteins.";
RL   J. Biol. Chem. 266:19079-19087(1991).
SQ   SEQUENCE   8 AA;  846 MW;  C007272DD865BAAA CRC64;


  Query Match             31.1%;  Score 14;  DB 13;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 SGA 5
            |||
Db        3 SGA 5


RESULT   8
Q9R7E8
ID   Q9R7E8      PRELIMINARY;      PRT;     9 AA.
AC   Q9R7E8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   KPSD PROTEIN (FRAGMENT).
GN   KPSD.
OS   Escherichia coli.
OG   Plasmid pCR3.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95180691.
RA   Rosenow C., Roberts I.S., Jann K.;
RT   "Isolation from recombinant Escherichia coli and characterization of
RT   CMP-Kdo synthetase, involved in the expression of the capsular K5
RT   polysaccharide (K-CKS).";
RL   FEMS Microbiol. Lett. 125:159-164(1995).
DR   EMBL; S76943; CAB33515.1; -.
KW   Plasmid.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  899 MW;  3EBBB72042C33DD8 CRC64;


  Query Match             31.1%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        4 GANINL 9
            || : |
Db        2 GAKVIL 7


RESULT   9
Q9TRS0
ID   Q9TRS0      PRELIMINARY;      PRT;     9 AA.
AC   Q9TRS0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CALCYCLIN-ASSOCIATED PROTEIN, CAP50=CA2+/PHOSPHOLIPID-BINDING PROTEIN
DE   L-7 FRAGMENT.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92250478.
RA   Tokumitsu H., Mizutani A., Minami H., Kobayashi R., Hidaka H.;
RT   "A calcyclin-associated protein is a newly identified member of the
RT   Ca2+/phospholipid-binding proteins, annexin family.";
RL   J. Biol. Chem. 267:8919-8924(1992).
SQ   SEQUENCE   9 AA;  1010 MW;  64E419C44865B72B CRC64;


  Query Match             31.1%;  Score 14;  DB 6;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSG 4
            |||
Db        3 LSG 5


RESULT   10
Q31415
ID   Q31415      PRELIMINARY;      PRT;     9 AA.
AC   Q31415;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
```

```
DE   MHC CLASS I ANTIGEN (FRAGMENT).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kanki T., Kuwasawa N., Sekiya Y., Ichikawa Y.;
RT   "Responsive expression of a MHC class I epitope and genes following
RT   Marek's disease virus infection.";
RL   Submitted (JAN-1992) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; D90399; BAA14395.1; -.
KW   MHC.
FT   NON_TER      1      1
SQ   SEQUENCE   9 AA;  859 MW;  8A55A76455B861B5 CRC64;


  Query Match             31.1%;  Score 14;  DB 7;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 SGAN 6
            :|:|
Db        3 TGSN 6


RESULT  11
O35953
ID   O35953     PRELIMINARY;      PRT;     9 AA.
AC   O35953;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   SODIUM CHANNEL, VOLTAGE-GATED, TYPE VIII, ALPHA POLYPEPTIDE
DE   (TRUNCATED VOLTAGE-GATED SODIUM CHANNEL ALPHA SUBUNIT) (FRAGMENT).
GN   SCN8A.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIII;
RX   MEDLINE; 97442476.
RA   Plummer N.W., McBurney M.W., Meisler M.H.;
RT   "Alternative splicing of the sodium channel SCN8A predicts a truncated
RT   two-domain protein in fetal brain and non-neuronal cells.";
RL   J. Biol. Chem. 272:24008-24015(1997).
DR   EMBL; U97672; AAB80914.1; -.
DR   MGD; MGI:103169; Scn8a.
KW   Ionic channel.
FT   NON_TER      1      1
SQ   SEQUENCE   9 AA;  898 MW;  22D92865B735B737 CRC64;

  Query Match             31.1%;  Score 14;  DB 11;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSG 4
            |||
Db        5 LSG 7


RESULT  12
Q89491
ID   Q89491     PRELIMINARY;      PRT;     9 AA.
AC   Q89491;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 1.1 KDA PROTEIN.
OS   Murine minute virus (Murine parvovirus).
OC   Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LYMPHOTROPIC VARIANT;
RX   MEDLINE; 86115415.
RA   Astell C.R., Gardiner E.M., Tattersall P.;
RT   "DNA sequence of the lymphotropic variant of minute virus of mice,
RT   MVM(i), and comparison with the DNA sequence of the fibrotropic
RT   prototype strain.";
RL   J. Virol. 570:656-669(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MVM(P);
RX   MEDLINE; 83143341.
RA   Astell C.R., Thomson M., Merchlinsky M., Ward D.C.;
RT   "The complete DNA sequence of minute virus of mice, an autonomous
RT   parvovirus.";
RL   Nucleic Acids Res. 11:999-1018(1983).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MVM(P);
RX   MEDLINE; 86115415.
RA   Astell C.R., Gardiner E.M., Tattersall P.;
RT   "DNA sequence of the lymphotropic variant of minute virus of mice,
RT   MVM(i), and comparison with the DNA sequence of the fibrotropic
RT   prototype strain.";
RL   J. Virol. 57:656-669(1986).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MVM(P);
RX   MEDLINE; 87061199.
RA   Morgan W.R., Ward D.C.;
RT   "Three splicing patterns are used to excise the small intron common to
RT   all minute virus of mice RNAs.";
RL   J. Virol. 60:1170-1174(1986).
DR   EMBL; M12032; AAA69570.1; -.
DR   EMBL; J02275; AAA67112.1; -.
DR   EMBL; V01115; CAA24311.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   9 AA;  1061 MW;  C3FD405863637862 CRC64;

  Query Match             31.1%;  Score 14;  DB 12;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 GANI 7
            | |:
Db        6 GINV 9


RESULT  13
Q07624
ID   Q07624     PRELIMINARY;      PRT;     7 AA.
AC   Q07624;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL PROTEIN (FRAGMENT).
OS   Rous sarcoma virus.
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PRAGUE C;
RX   MEDLINE; 93010967.
RA   Donze O., Spahr P.F.;
RT   "Role of the open reading frames of Rous sarcoma virus leader RNA in
RT   translation and genome packaging.";
RL   EMBO J. 11:3747-3757(1992).
DR   EMBL; X67587; CAA47862.1; -.
KW   Hypothetical protein.
FT   NON_TER      7      7
SQ   SEQUENCE   7 AA;  672 MW;  776045A7687DD6F0 CRC64;
```

```
  Query Match             28.9%;  Score 13;  DB 12;  Length 7;
  Best Local Similarity   33.3%;  Pred. No. 3e+05;
  Matches    2;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LSGANI 7
           ::| :|
Db       1 MAGPSI 6

RESULT  14
Q90498
ID   Q90498      PRELIMINARY;      PRT;      8 AA.
AC   Q90498;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   MYOGLOBIN (FRAGMENT).
OS   Erythrura gouldiae (Gouldian finch).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Passeridae; Erythrura.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EGG1;
RX   MEDLINE; 98208049.
RA   Heslewood M.M., Elphinstone M.S., Tidemann S.C., Baverstock P.R.;
RT   "Myoglobin intron variation in the Gouldian Finch Erythrura gouldiae
RT   assessed by temperature gradient gel electrophoresis.";
RL   Electrophoresis 19:142-151(1998).
DR   EMBL; U40496; AAC60363.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  890 MW;  C6C1F2C865B046DE CRC64;

  Query Match             28.9%;  Score 13;  DB 13;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LSGAN 6
           :|| :
Db       3 ISGVH 7

RESULT  15
Q91098
ID   Q91098      PRELIMINARY;      PRT;      8 AA.
AC   Q91098;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   MYOGLOBIN (FRAGMENT).
OS   Manorina melanocephala (noisy miner).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Meliphagidae; Manorina.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D02;
RX   MEDLINE; 98208049.
RA   Heslewood M.M., Elphinstone M.S., Tidemann S.C., Baverstock P.R.;
RT   "Myoglobin intron variation in the Gouldian Finch Erythrura gouldiae
RT   assessed by temperature gradient gel electrophoresis.";
RL   Electrophoresis 19:142-151(1998).
DR   EMBL; U40497; AAC60364.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  890 MW;  C6C1F2C865B046DE CRC64;

  Query Match             28.9%;  Score 13;  DB 13;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LSGAN 6
           :|| :
Db       3 ISGVH 7

RESULT  16
Q48686
ID   Q48686      PRELIMINARY;      PRT;      9 AA.
AC   Q48686;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   PROMOTER 23 DNA FRAGMENT (FRAGMENT).
OS   Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Lactococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88105390.
RA   van der Vossen J.M., der Lelie D., Venema G.;
RT   "Isolation and characterization of Streptococcus cremoris Wg2-specific
RT   promoters.";
RL   Appl. Environ. Microbiol. 53:2452-2457(1987).
DR   EMBL; M24763; AAA74720.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1080 MW;  5AF3A44AA4469443 CRC64;

  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 NIN 8
           |:|
Db       3 NMN 5

RESULT  17
Q44001
ID   Q44001      PRELIMINARY;      PRT;      9 AA.
AC   Q44001;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (STRAIN CDC 9179-79) (FRAGMENT).
GN   EXEF'.
OS   Aeromonas eucrenophila.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 9179-79;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89461; CAA61637.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1059 MW;  785852C69444472B CRC64;

  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 INL 9
           :||
Db       7 VNL 9

RESULT  18
Q44377
```

```
ID   Q44377      PRELIMINARY;      PRT;     9 AA.
AC   Q44377;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (STRAIN ATCC 49659) (FRAGMENT).
GN   EXEF'.
OS   Aeromonas trota.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 49659;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89468; CAA61651.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1059 MW;  785852C69444472B CRC64;

  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 INL 9
           :||
Db       7 VNL 9

RESULT  19
Q44468
ID   Q44468      PRELIMINARY;      PRT;     9 AA.
AC   Q44468;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (STRAIN CDC 1306-83) (FRAGMENT).
GN   EXEF'.
OS   Aeromonas veronii.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1306-83;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89457; CAA61629.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1059 MW;  785852C69444472B CRC64;

  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 INL 9
           :||
Db       7 VNL 9

RESULT  20
Q43928
ID   Q43928      PRELIMINARY;      PRT;     9 AA.
AC   Q43928; Q43918; Q43920; Q43921;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
```

```
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (FRAGMENT).
GN   EXEF'.
OS   Aeromonas caviae.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VARIOUS STRAINS;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89464; CAA61643.1; -.
DR   EMBL; X89462; CAA61639.1; -.
DR   EMBL; X89460; CAA61635.1; -.
DR   EMBL; X89463; CAA61641.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1059 MW;  785852C69444472B CRC64;

  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 INL 9
           :||
Db       7 VNL 9

RESULT  21
O78337
ID   O78337      PRELIMINARY;      PRT;     9 AA.
AC   O78337;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa leprieurii.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=736;
RA   Kamiya M., West J.A., King R.J., Zuccarello G.C., Tanaka J., Hara Y.;
RT   "Evolutionary divergence in the red algae Caloglossa leprieurii and C.
RT   apomeiotica.";
RL   J. Phycol. 34:361-370(1998).
DR   EMBL; D87813; BAA31279.1; -.
KW   Chloroplast.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             28.9%;  Score 13;  DB 8;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ANI 7
           ||:
Db       7 ANV 9

RESULT  22
Q9TLD0
ID   Q9TLD0      PRELIMINARY;      PRT;     9 AA.
AC   Q9TLD0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
```

```
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Bostrychia moritziana.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC   Bostrychia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M3189.CO;
RA   Zuccarello G.C., West J.A., King R.J.;
RT   "Biogeography of Bostrychia moritziana (Ceramiales).";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF126705; AAD55863.1; -.
KW   Chloroplast.
FT   NON_TER      1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             28.9%;  Score 13;  DB 8;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ANI 7
           ||:
Db       7 ANV 9

RESULT  23
Q9TJ87
ID   Q9TJ87      PRELIMINARY;      PRT;     9 AA.
AC   Q9TJ87;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa monosticha.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=987;
RA   Kamiya M., Tanaka J., King R.J., West J.A., Zuccarello G.C., Kawai H.;
RT   "Reproductive and genetic distinction between broad and narrow
RT   entities of Caloglossa continua (Delesseriaceae, Rhodophyta).";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB023381; BAA88912.1; -.
KW   Chloroplast.
FT   NON_TER      1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             28.9%;  Score 13;  DB 8;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ANI 7
           ||:
Db       7 ANV 9

RESULT  24
Q9TJ85
ID   Q9TJ85      PRELIMINARY;      PRT;     9 AA.
AC   Q9TJ85;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
```

```
GN   RBCL.
OS   Caloglossa stipitata.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Kamiya M., Tanaka J., King R.J., West J.A., Zuccarello G.C., Kawai H.;
RT   "Reproductive and genetic distinction between broad and narrow
RT   entities of Caloglossa continua (Delesseriaceae, Rhodophyta).";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB023384; BAA88918.1; -.
KW   Chloroplast.
FT   NON_TER      1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             28.9%;  Score 13;  DB 8;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ANI 7
           ||:
Db       7 ANV 9

RESULT  25
Q9T389
ID   Q9T389      PRELIMINARY;      PRT;     9 AA.
AC   Q9T389;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa postiae.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAW1122, AND 962;
RA   Kamiya M., Tanaka J., King R.J., West J.A., Zuccarello G.C., Kawai H.;
RT   "Reproductive and genetic distinction between broad and narrow
RT   entities of Caloglossa continua (Delesseriaceae, Rhodophyta).";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB023383; BAA88916.1; -.
DR   EMBL; AB023382; BAA88914.1; -.
KW   Chloroplast.
FT   NON_TER      1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             28.9%;  Score 13;  DB 8;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ANI 7
           ||:
Db       7 ANV 9

RESULT  26
Q9T388
ID   Q9T388      PRELIMINARY;      PRT;     9 AA.
AC   Q9T388;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
```

```
GN   RBCL.
OS   Caloglossa continua.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=729, AND 500;
RA   Kamiya M., Tanaka J., King R.J., West J.A., Zuccarello G.C., Kawai H.;
RT   "Reproductive and genetic distinction between broad and narrow
RT   entities of Caloglossa continua (Delesseriaceae, Rhodophyta).";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB023380; BAA88910.1; -.
DR   EMBL; AB023379; BAA88908.1; -.
KW   Chloroplast.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             28.9%;  Score 13;  DB 8;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ANI 7
           ||:
Db       7 ANV 9

RESULT  27
Q9T387
ID   Q9T387      PRELIMINARY;      PRT;     9 AA.
AC   Q9T387;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Bostrychia radicans.
OG   Chloroplast.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC   Bostrychia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R.3881.QLD, M3001.MI, R3826.BZ, AND B.R.F.M.BZ;
RA   Zuccarello G.C., West J.A., King R.J.;
RT   "Biogeography of Bostrychia moritziana (Ceramiales).";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF126715; AAD55883.1; -.
DR   EMBL; AF126701; AAD55855.1; -.
DR   EMBL; AF126704; AAD55861.1; -.
DR   EMBL; AF126706; AAD55865.1; -.
KW   Chloroplast.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             28.9%;  Score 13;  DB 8;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ANI 7
           ||:
Db       7 ANV 9

RESULT  28
O81962
ID   O81962      PRELIMINARY;      PRT;     9 AA.
AC   O81962;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa apomeiotica.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=910;
RA   Kamiya M., West J.A., King R.J., Zuccarello G.C., Tanaka J., Hara Y.;
RT   "Evolutionary divergence in the red algae Caloglossa leprieurii and C.
RT   apomeiotica.";
RL   J. Phycol. 34:361-370(1998).
DR   EMBL; D89948; BAA31281.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             28.9%;  Score 13;  DB 10;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ANI 7
           ||:
Db       7 ANV 9

RESULT  29
O81964
ID   O81964      PRELIMINARY;      PRT;     9 AA.
AC   O81964;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa continua.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=639;
RA   Kamiya M., West J.A., King R.J., Zuccarello G.C., Tanaka J., Hara Y.;
RT   "Evolutionary divergence in the red algae Caloglossa leprieurii and C.
RT   apomeiotica.";
RL   J. Phycol. 34:361-370(1998).
DR   EMBL; D89950; BAA31285.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             28.9%;  Score 13;  DB 10;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ANI 7
           ||:
Db       7 ANV 9

RESULT  30
O81966
ID   O81966      PRELIMINARY;      PRT;     9 AA.
AC   O81966;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa monosticha.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
```

```
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=892;
RA   Kamiya M., West J.A., King R.J., Zuccarello G.C., Tanaka J., Hara Y.;
RT   "Evolutionary divergence in the red algae Caloglossa leprieurii and C.
RT   apomeiotica.";
RL   J. Phycol. 34:361-370(1998).
DR   EMBL; D89960; BAA31305.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             28.9%;  Score 13;  DB 10;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ANI 7
           ||:
Db       7 ANV 9


RESULT  31
O81968
ID   O81968      PRELIMINARY;      PRT;     9 AA.
AC   O81968;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa ogasawaraensis.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=596;
RA   Kamiya M., West J.A., King R.J., Zuccarello G.C., Tanaka J., Hara Y.;
RT   "Evolutionary divergence in the red algae Caloglossa leprieurii and C.
RT   apomeiotica.";
RL   J. Phycol. 34:361-370(1998).
DR   EMBL; D89961; BAA31307.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             28.9%;  Score 13;  DB 10;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ANI 7
           ||:
Db       7 ANV 9


RESULT  32
O82778
ID   O82778      PRELIMINARY;      PRT;     9 AA.
AC   O82778;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT
DE   (FRAGMENT).
GN   RBCL.
OS   Caloglossa leprieurii.
OC   Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Delesseriaceae;
OC   Caloglossa.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1048, 902, 490, 932, 922, 880, 1053, 1052;
RA   Kamiya M., West J.A., King R.J., Zuccarello G.C., Tanaka J., Hara Y.;
```

```
RT   "Evolutionary divergence in the red algae Caloglossa leprieurii and C.
RT   apomeiotica.";
RL   J. Phycol. 34:361-370(1998).
DR   EMBL; D89959; BAA31303.1; -.
DR   EMBL; D89949; BAA31283.1; -.
DR   EMBL; D89951; BAA31287.1; -.
DR   EMBL; D89952; BAA31289.1; -.
DR   EMBL; D89953; BAA31291.1; -.
DR   EMBL; D89954; BAA31293.1; -.
DR   EMBL; D89955; BAA31295.1; -.
DR   EMBL; D89956; BAA31297.1; -.
DR   EMBL; D89957; BAA31299.1; -.
DR   EMBL; D89958; BAA31301.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  977 MW;  CA1A4DC1B771AB02 CRC64;

  Query Match             28.9%;  Score 13;  DB 10;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ANI 7
           ||:
Db       7 ANV 9


RESULT  33
Q15901
ID   Q15901      PRELIMINARY;      PRT;     8 AA.
AC   Q15901;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   (CLONE XP7B11B) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   Lee C.C., Yazdani A., Wehnert M., Bailey J., Couch L., Xiong M.,
RA   Coolbaugh M.I., Chinault C.A., Baldini A., Lindsay E.A., Zhao Z.Y.,
RA   Caskey C.T.H.;
RL   Hum. Mol. Genet. 0:0-0(0).
DR   EMBL; L32080; AAA73891.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  860 MW;  37D72878676729CB CRC64;

  Query Match             26.7%;  Score 12;  DB 4;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           :| |
Db       2 FLPG 5


RESULT  34
Q9UMH9
ID   Q9UMH9      PRELIMINARY;      PRT;     8 AA.
AC   Q9UMH9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RHCE PROTEIN (FRAGMENT).
GN   RHCE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   TISSUE=BLOOD;
RA   Matassi G., Cherif-Zahar B., Mouro I., Cartron J.P.;
RT   "Characterization of the recombination hot spot involved in the
RT   genomic rearrangement leading to the hybrid D-CE-D gene in the DVI
RT   phenotype.";
RL   Am. J. Hum. Genet. 60:808-817(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE; 90349591.
RA   Cherif-Zahar B., Bloy C., Le Van Kim C., Blanchard D., Bailly P.,
RA   Hermand P., Salmon C., Cartron J.-P., Colin Y.;
RT   "Molecular cloning and protein structure of a human blood group Rh
RT   polypeptide.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:6243-6247(1990).
DR   EMBL; Z97030; CAB09726.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  1049 MW;  C007244691FB5AB1 CRC64;


  Query Match             26.7%;  Score 12;  DB 4;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 NINL 9
           ::||
Db       4 HMNL 7


RESULT  35
O02032
ID   O02032      PRELIMINARY;      PRT;     8 AA.
AC   O02032;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   METALLOTHIONEIN (FRAGMENT).
GN   LPMT2.
OS   Lytechinus pictus (Painted sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
OC   Euechinoidea; Echinacea; Temnopleuroida; Toxopneustidae; Lytechinus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97264487.
RA   Cserjesi P., Fang H., Brandhorst B.P.;
RT   "Metallothionein gene expression in embryos of the sea urchin
RT   Lytechinus pictus.";
RL   Mol. Reprod. Dev. 47:39-46(1997).
DR   EMBL; U83400; AAB58320.1; -.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  823 MW;  EBD5A2C1F7686766 CRC64;


  Query Match             26.7%;  Score 12;  DB 5;  Length 8;
  Best Local Similarity   12.5%;  Pred. No. 3e+05;
  Matches    1;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LSGANINL 9
           : | ::::
Db       1 MPGPHVSV 8


Search completed: December 16, 2000, 04:22:13
Job time: 4607 sec
```

```
                        GenCore version 4.5
              Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 04:23:31 ; Search time 68.03 Seconds
                                               (without alignments)
                                               4.227 Million cell updates/sec

Title:           US-09-529-121-4
Perfect score:   45
Sequence:        1 YLSGANINL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters:      213

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       SwissProt_39:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                  %
Result           Query
  No.   Score   Match  Length DB  ID                          Description
-------------------------------------------------------------------------------
    1     17     37.8      8  1   CPD1_ENTFA                  P13269 enterococcu
    2     16     35.6      8  1   FAR8_CALVO                  P41863 calliphora
    3     15     33.3      9  1   ISOT_CYPCA                  P42993 cyprinus ca
    4     13     28.9      9  1   OXYT_RAJCL                  P42994 raja clavat
    5     13     28.9      9  1   THYF_PIG                    P01255 sus scrofa
    6     12     26.7      7  1   IGAO_DACDE                  P06294 dactylium d
    7     12     26.7      9  1   PGLR_DIAAB                  P81179 diaprepes a
    8     11     24.4      4  1   FAR3_HIRME                  P42562 hirudo medi
    9     11     24.4      5  1   PRCT_PERAM                  P01373 periplaneta
   10     11     24.4      7  1   FAR2_ASCSU                  P31890 ascaris suu
   11     11     24.4      7  1   GFRP_MOUSE                  P99025 mus musculu
   12     11     24.4      7  1   LANC_CARUI                  P36960 carnobacter
   13     11     24.4      8  1   CAD1_ENTFA                  P13268 enterococcu
   14     11     24.4      8  1   LCK8_LEUMA                  P19990 leucophaea
   15     11     24.4      9  1   MOSF_CLYJA                  P19853 clypeaster
   16     11     24.4      9  1   OXYA_SCYCA                  P42996 scyliorhinu
   17     11     24.4      9  1   OXYT_BUFRE                  P42995 bufo regula
   18     11     24.4      9  1   TAL3_PICJA                  P17441 pichia jadi
   19     10     22.2      5  1   TPIS_CANFA                  P54714 canis famil
   20     10     22.2      5  1   UXA4_CHLTR                  P38005 chlamydia t
   21     10     22.2      6  1   CIP2_MYTED                  P13737 mytilus edu
   22     10     22.2      6  1   TMOF_SARBU                  P41495 sarcophaga
   23     10     22.2      7  1   ALL2_CARMA                  P81805 carcinus ma
   24     10     22.2      7  1   ALL3_CARMA                  P81806 carcinus ma
   25     10     22.2      7  1   ALL4_CARMA                  P81807 carcinus ma
   26     10     22.2      7  1   ALL5_CARMA                  P81808 carcinus ma
   27     10     22.2      7  1   UN06_PINPS                  P81675 pinus pinas
   28     10     22.2      8  1   AL12_CARMA                  P81815 carcinus ma
   29     10     22.2      8  1   AL17_CARMA                  P81820 carcinus ma
   30     10     22.2      8  1   ALL1_CYDPO                  P82152 cydia pomon
   31     10     22.2      8  1   ALL6_CYDPO                  P82157 cydia pomon
   32     10     22.2      8  1   ALL7_CARMA                  P81809 carcinus ma
   33     10     22.2      8  1   ALL8_CARMA                  P81811 carcinus ma
   34     10     22.2      8  1   ALL9_CARMA                  P81812 carcinus ma
   35     10     22.2      8  1   B44K_PORGI                  P81886 porphyromon
   36     10     22.2      8  1   GLUR_HUMAN                  P02729 homo sapien
   37     10     22.2      8  1   LCK3_LEUMA                  P21142 leucophaea
   38     10     22.2      8  1   LCK5_LEUMA                  P19987 leucophaea
   39     10     22.2      8  1   LPMS_STAEP                  P23211 staphylococ
   40     10     22.2      8  1   PLP_BRANA                   P81707 brassica na
   41     10     22.2      8  1   UF06_MOUSE                  P38644 mus musculu
   42     10     22.2      8  1   UPA1_HUMAN                  P30087 homo sapien
   43     10     22.2      8  1   VGLG_HSV2B                  P81780 herpes simp
   44     10     22.2      9  1   AL10_CARMA                  P81813 carcinus ma
   45     10     22.2      9  1   AL11_CARMA                  P81814 carcinus ma
   46     10     22.2      9  1   BUK_CLOPA                   P81337 clostridium
   47     10     22.2      9  1   DSIP_RABIT                  P01158 oryctolagus
   48     10     22.2      9  1   FAR5_ASCSU                  P43170 ascaris suu
   49     10     22.2      9  1   FAR6_CALVO                  P41861 calliphora
   50     10     22.2      9  1   FARP_CALSI                  P38495 callinectes
   51     10     22.2      9  1   FIBB_MACFU                  P19345 macaca fusc
   52     10     22.2      9  1   MOSH_CLYJA                  P19852 clypeaster
   53     10     22.2      9  1   OXYA_SQUAC                  P42999 squalus aca
   54     10     22.2      9  1   OXYV_SQUAC                  P43000 squalus aca
   55     10     22.2      9  1   TKL1_LOCMI                  P16223 locusta mig
   56     10     22.2      9  1   TRP4_LEUMA                  P81736 leucophaea
   57     10     22.2      9  1   ULAH_HUMAN                  P31934 homo sapien
   58      9     20.0      4  1   FAR4_HIRME                  P42563 hirudo medi
   59      9     20.0      5  1   AL14_CARMA                  P81817 carcinus ma
   60      9     20.0      5  1   FARP_ARTTR                  P41853 artioposthi
   61      9     20.0      7  1   CHOX_ALCSP                  P16101 alcaligenes
   62      9     20.0      7  1   FAR1_PROCL                  P38499 procambarus
   63      9     20.0      7  1   FAR3_HAECO                  P81298 haemonchus
   64      9     20.0      7  1   FAR3_PANRE                  P41874 panagrellus
   65      9     20.0      7  1   FAR5_HIRME                  P42564 hirudo medi
   66      9     20.0      8  1   AKHG_GRYBI                  P14086 gryllus bim
   67      9     20.0      8  1   AKH_LIBAU                   P25418 libellula a
   68      9     20.0      8  1   AL15_CARMA                  P81818 carcinus ma
   69      9     20.0      8  1   AL16_CARMA                  P81819 carcinus ma
   70      9     20.0      8  1   AL18_CARMA                  P81821 carcinus ma
   71      9     20.0      8  1   ALL3_CYDPO                  P82154 cydia pomon
   72      9     20.0      8  1   ALL4_CALVO                  P41840 calliphora
   73      9     20.0      8  1   ALL4_CYDPO                  P82155 cydia pomon
   74      9     20.0      8  1   ALL5_CYDPO                  P82156 cydia pomon
   75      9     20.0      8  1   ANG2_BOTJA                  Q10582 bothrops ja

                               ALIGNMENTS


RESULT   1
CPD1_ENTFA
ID   CPD1_ENTFA     STANDARD;      PRT;     8 AA.
AC   P13269;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   SEX PHEROMONE CPD1.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 85040388.
RA   Suzuki A., Mori M., Sagakami Y., Isogai A., Fujino M., Kitada C.,
RA   Craig R.A., Clewell D.B.;
RT   "Isolation and structure of bacterial sex pheromone, cPD1.";
RL   Science 226:849-850(1984).
CC   -!- FUNCTION: CPD1 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE
CC       BACTERIOCIN PLASMID PPD1.
KW   Pheromone.
SQ   SEQUENCE   8 AA;  913 MW;  8665B729C682C729 CRC64;

  Query Match             37.8%;  Score 17;  DB 1;  Length 8;
  Best Local Similarity   75.0%;  Pred. No. 8.8e+04;
```

```
  Matches     3;  Conservative     1;  Mismatches     0;  Indels     0;  Gaps     0;

Qy         1 YLSG 4
             :|||
Db         5 FLSG 8


RESULT   2
FAR8_CALVO
ID   FAR8_CALVO     STANDARD;      PRT;     8 AA.
AC   P41863;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 8.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; H41978; H41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  957 MW;  72D40699CAA44DD8 CRC64;


  Query Match             35.6%;  Score 16;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches     3;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy         4 GAN 6
             |||
Db         1 GAN 3


RESULT   3
ISOT_CYPCA
ID   ISOT_CYPCA     STANDARD;      PRT;     9 AA.
AC   P42993;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ISOTOCIN.
OS   Cyprinus carpio (Common carp).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PITUITARY;
RA   Acher R., Chauvet J., Chauvet M.-T., Crepy D.;
RT   "Characterization of neurohypophyseal hormones from a fresh water bony
RT   fish, the carp (Cyprinus carpio). Comparison with hormones from sea
RT   water bony fishs.";
RL   Comp. Biochem. Physiol. 14:245-254(1965).
CC   -!- FUNCTION: ANTIDIURETIC HORMONE.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   PIR; A61364; A61364.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
```

```
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  969 MW;  17FF476EB455B04B CRC64;

  Query Match             33.3%;  Score 15;  DB 1;  Length 9;
  Best Local Similarity   42.9%;  Pred. No. 8.8e+04;
  Matches     3;  Conservative     1;  Mismatches     3;  Indels     0;  Gaps     0;

Qy         1 YLSGANI 7
             |:|   |
Db         2 YISNCPI 8


RESULT   4
OXYT_RAJCL
ID   OXYT_RAJCL     STANDARD;      PRT;     9 AA.
AC   P42994;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   GLUMITOCIN.
OS   Raja clavata (Thornback ray).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Hypnosqualea; Pristiorajea; Batoidea;
OC   Rajiformes; Rajidae; Raja.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 66123415.
RA   Acher R., Chauvet J., Chauvet M.-T., Crepy D.;
RT   "Phylogeny of neurophyophyseal peptides: isolation of a new hormone,
RT   glumitocin (Ser 4-Gln 8-ocytocin) present in a cartilaginous fish,
RT   the ray (Raia clavata).";
RL   Biochim. Biophys. Acta 107:393-396(1965).
CC   -!- FUNCTION: ANTIDIURETIC HORMONE.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  984 MW;  17E9C76EB455B04B CRC64;


  Query Match             28.9%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches     2;  Conservative     1;  Mismatches     0;  Indels     0;  Gaps     0;

Qy         1 YLS 3
             |:|
Db         2 YIS 4


RESULT   5
THYF_PIG
ID   THYF_PIG       STANDARD;      PRT;     9 AA.
AC   P01255;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   21-JUL-1986 (Rel. 01, Last annotation update)
DE   THYMIC FACTOR.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 78026571.
RA   Pleau J.-M., Dardenne M., Blouquit Y., Bach J.-F.;
RT   "Structural study of circulating thymic factor: a peptide isolated
RT   from pig serum. II. Amino acid sequence.";
RL   J. Biol. Chem. 252:8045-8047(1977).
CC   -!- MISCELLANEOUS: THE BIOLOGICAL SOURCE(S) AND PHYSIOLOGICAL
```

```
CC         ACTIVITIES OF THIS PEPTIDE HAVE NOT BEEN DETERMINED.
DR    PIR; A01523; YFPG.
FT    MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
SQ    SEQUENCE   9 AA;  876 MW;  D500B87866C5B33D CRC64;


  Query Match             28.9%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GAN 6
           |:|
Db       7 GSN 9


RESULT   6
IGAO_DACDE
ID    IGAO_DACDE     STANDARD;      PRT;     7 AA.
AC    P06294;
DT    01-JAN-1988 (Rel. 06, Created)
DT    01-JAN-1988 (Rel. 06, Last sequence update)
DT    01-OCT-1994 (Rel. 30, Last annotation update)
DE    GALACTOSE OXIDASE INHIBITOR.
OS    Dactylium dendroides (Cladobotryum dendroides).
OC    Eukaryota; Fungi; Ascomycota; Hypocreales; Hypocreaceae; Hypomyces.
RN    [1]
RP    SEQUENCE.
RA    Avigad G., Markus Z.;
RT    "Identification of a peptide inhibitor of galactose oxidase from
RT    Dactylium dendroides.";
RL    Fed. Proc. 31:447-447(1972).
CC    -!- FUNCTION: BINDS ONE COPPER ION PER MOLECULE BUT DOES NOT BIND THE
CC        GALACTOSE OXIDASE APOENZYME. IT MAY INACTIVATE THE ENZYME BY
CC        BINDING TO ITS PROSTHETIC COPPER GROUP.
DR    PIR; A01341; XEYDGD.
KW    Copper; Metalloenzyme inhibitor.
SQ    SEQUENCE   7 AA;  706 MW;  75BB01A456D87DB0 CRC64;


  Query Match             26.7%;  Score 12;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAN 6
            :| |
Db       1 AGQN 4


RESULT   7
PGLR_DIAAB
ID    PGLR_DIAAB     STANDARD;      PRT;     9 AA.
AC    P81179;
DT    15-JUL-1998 (Rel. 36, Created)
DT    15-JUL-1998 (Rel. 36, Last sequence update)
DT    30-MAY-2000 (Rel. 39, Last annotation update)
DE    ENDO-POLYGALACTURONASE (PG) (EC 3.2.1.15) (FRAGMENT).
OS    Diaprepes abbreviatus (Sugarcane rootstalk borer weevil).
OC    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC    Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC    Cucujiformia; Curculionidae; Entiminae; Entimini; Diaprepes.
RN    [1]
RP    SEQUENCE.
RC    TISSUE=LARVAL GUT;
RA    Doostdar H., McCollum T.G., Mayer R.T.;
RT    "Purification and characterization of an endo-polygalacturonase from
RT    the gut of West Indies sugarcane rootstalk borer weevil (Diaprepes
RT    abbreviatus L.) larvae.";
RL    Comp. Biochem. Physiol. 118B:861-867(1997).
CC    -!- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-ALPHA-D-
CC        GALACTOSIDURONIC LINKAGES IN PECTATE AND OTHER GALACTURONANS.
CC    -!- INDUCTION: INHIBITED BY CITRUS PGIP.
CC    -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS PROTEIN IS:
CC        9.4, ITS MW IS: 44.5 KDA.
```

```
CC    -!- SIMILARITY: WEAK TO OTHER POLYGALACTURONASES.
KW    Hydrolase; Glycosidase; Cell wall.
FT    NON_TER       9      9
SQ    SEQUENCE   9 AA;  1041 MW;  1F49087042DB41BB CRC64;


  Query Match             26.7%;  Score 12;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |: |
Db       4 YVIG 7


RESULT   8
FAR3_HIRME
ID    FAR3_HIRME     STANDARD;      PRT;     4 AA.
AC    P42562;
DT    01-NOV-1995 (Rel. 32, Created)
DT    01-NOV-1995 (Rel. 32, Last sequence update)
DT    01-NOV-1995 (Rel. 32, Last annotation update)
DE    FMRFAMIDE-LIKE NEUROPEPTIDE YLRF-AMIDE.
OS    Hirudo medicinalis (Medicinal leech).
OC    Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;
OC    Arynchobdellida; Hirudiniformes; Hirudinidae; Hirudo.
RN    [1]
RP    SEQUENCE.
RX    MEDLINE; 92195954.
RA    Evans B.D., Pohl J., Kartsonis M.A., Calabrese R.L.;
RT    "Identification of RFamide neuropeptides in the medicinal leech.";
RL    Peptides 12:897-908(1991).
CC    -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC        FAMILY.
KW    Neuropeptide; Amidation.
FT    MOD_RES       4      4       AMIDATION.
SQ    SEQUENCE   4 AA;  598 MW;  69D4073B30000000 CRC64;


  Query Match             24.4%;  Score 11;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YL 2
           ||
Db       1 YL 2


RESULT   9
PRCT_PERAM
ID    PRCT_PERAM     STANDARD;      PRT;     5 AA.
AC    P01373;
DT    21-JUL-1986 (Rel. 01, Created)
DT    21-JUL-1986 (Rel. 01, Last sequence update)
DT    01-FEB-1995 (Rel. 31, Last annotation update)
DE    PROCTOLIN.
OS    Periplaneta americana (American cockroach),
OS    Limulus polyphemus (Atlantic horseshoe crab), and
OS    Carcinus maenas (Common shore crab) (Green crab).
OC    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC    Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC    Blattoidea; Blattidae; Periplaneta.
RN    [1]
RP    SEQUENCE.
RC    SPECIES=P.AMERICANA;
RX    MEDLINE; 76074708.
RA    Starratt A.N., Brown B.E.;
RT    "Structure of the pentapeptide proctolin, a proposed neurotransmitter
RT    in insects.";
RL    Life Sci. 17:1253-1256(1975).
RN    [2]
RP    BIOLOGICAL SOURCE.
RC    SPECIES=P.AMERICANA;
```

```
RX   MEDLINE; 81225865.
RA   O'Shea M., Adams M.E.;
RT   "Pentapeptide (proctolin) associated with an identified neuron.";
RL   Science 213:567-569(1981).
RN   [3]
RP   SEQUENCE.
RC   SPECIES=L.POLYPHEMUS;
RX   MEDLINE; 90287800.
RA   Groome J.R., Tillinghast E.K., Townley M.A., Vetrovs A.,
RA   Watson W.H. III, Hunt D.F., Griffin P.R., Alexander J.E.,
RA   Shabanowitz J.;
RT   "Identification of proctolin in the central nervous system of the
RT   horseshoe crab, Limulus polyphemus.";
RL   Peptides 11:205-211(1990).
RN   [4]
RP   SEQUENCE.
RC   SPECIES=C.MAENAS;
RX   MEDLINE; 86232789.
RA   Stangier J., Dircksen H., Keller R.;
RT   "Identification and immunocytochemical localization of proctolin in
RT   pericardial organs of the shore crab, Carcinus maenas.";
RL   Peptides 7:67-72(1986).
CC   -!- FUNCTION: STIMULATES CARDIAC OUTPUT AND HINDGUT MOTILITY,
CC       MODULATES VISCERAL AND SKELETAL MUSCLE IN MANY ARTHROPODS.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE LATERAL WHITE NEURONS AND IN
CC       THE CRAB PERICARDIAL ORGANS.
DR   PIR; A01644; HOROHA.
DR   PIR; A60411; A60411.
KW   Neuropeptide.
SQ   SEQUENCE   5 AA;  649 MW;  71B7673B44600000 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YL 2
              ||
Db          2 YL 3


RESULT  10
FAR2_ASCSU
ID   FAR2_ASCSU     STANDARD;      PRT;     7 AA.
AC   P31890;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE AF2.
OS   Ascaris suum (Pig roundworm) (Ascaris lumbricoides), and
OS   Panagrellus redivivus.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC   Ascarididae; Ascaris.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=A.SUUM;
RX   MEDLINE; 93324431.
RA   Cowden C., Stretton A.O.W.;
RT   "AF2, an Ascaris neuropeptide: isolation, sequence, and bioactivity.";
RL   Peptides 14:423-430(1993).
RN   [2]
RP   SEQUENCE.
RC   SPECIES=P.REDIVIVUS;
RX   MEDLINE; 95060998.
RA   Maule A.G., Shaw C., Bowman J.W.;
RT   "The FMRFamide-like neuropeptide AF2 (Ascaris suum) is present in the
RT   free-living nematode, Panagrellus redivivus (Nematoda, Rhabditida).";
RL   Parasitology 109:351-356(1994).
CC   -!- FUNCTION: HAS EFFECTS ON MUSCLE TENSION.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE NERVE CORDS AND A VARIETY OF
CC       GANGLIA PARTICULARLY IN THE ANTERIOR REGIONS.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
```

```
KW   Neuropeptide; Amidation.
FT   MOD_RES       7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  992 MW;  69D4073B5B11E350 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YL 2
              ||
Db          4 YL 5


RESULT  11
GFRP_MOUSE
ID   GFRP_MOUSE     STANDARD;      PRT;     7 AA.
AC   P99025;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GTP CYCLOHYDROLASE I FEEDBACK REGULATORY PROTEIN (P35) (FRAGMENT).
GN   GFRP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LIVER;
RA   Sanchez J.-C., Rouge V., Frutiger S., Hughes G., Yan J.X.,
RA   Hoogland C., Appel R.D., Binz P.-A., Hochstrasser D.F.,
RA   Cowthorne M.;
RL   Submitted (AUG-1998) to the SWISS-PROT data bank.
CC   -!- FUNCTION: MEDIATES TETRAHYDROBIOPTERIN INHIBITION OF GTP
CC       CYCLOHYDROLASE I. THIS INHIBITION IS REVERSED BY L-PHENYLALANINE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
DR   SWISS-2DPAGE; P99025; MOUSE.
FT   INIT_MET      0      0
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  806 MW;  71B5B057273B4700 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YL 2
              ||
Db          2 YL 3


RESULT  12
LANC_CARUI
ID   LANC_CARUI     STANDARD;      PRT;     7 AA.
AC   P36960;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   LANTIBIOTIC CARNOCIN UI49 (FRAGMENT).
OS   Carnobacterium sp. (strain UI49).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC   Carnobacterium.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92321768.
RA   Stoffels G., Nissen-Meyer J., Gudmundsdottir A., Sletten K., Holo H.,
RA   Nes I.F.;
RT   "Purification and characterization of a new bacteriocin isolated from
RT   a Carnobacterium sp.";
RL   Appl. Environ. Microbiol. 58:1417-1422(1992).
CC   -!- FUNCTION: LANTHIONINE-CONTAINING PEPTIDE ANTIBIOTIC (LANTIBIOTIC).
CC       ACTIVE ON GRAM-POSITIVE BACTERIA.
```

```
KW   Antibiotic; Lantibiotic.
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  786 MW;  741776D05B05B810 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GANI 7
           |: |
Db       1 GSEI 4


RESULT  13
CAD1_ENTFA
ID   CAD1_ENTFA     STANDARD;      PRT;     8 AA.
AC   P13268;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   SEX PHEROMONE CAD1.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 85051889.
RA   Mori M., Sagakami Y., Narita M., Isogai A., Fujino M., Kitada C.,
RA   Craig R.A., Clewell D.B., Suzuki A.;
RT   "Isolation and structure of the bacterial sex pheromone, cAD1, that
RT   induces plasmid transfer in Streptococcus faecalis.";
RL   FEBS Lett. 178:97-100(1984).
CC   -!- FUNCTION: CAD1 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE
CC       HEMOLYSIN PLASMID PAD1.
KW   Pheromone.
SQ   SEQUENCE   8 AA;  819 MW;  047DD732C735B9C7 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           |:|
Db       6 LAG 8


RESULT  14
LCK8_LEUMA
ID   LCK8_LEUMA     STANDARD;      PRT;     8 AA.
AC   P19990;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   LEUCOKININ VIII (L-VIII).
OS   Leucophaea maderae (Madeira cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blaberoidea; Blaberidae; Leucophaea.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=HEAD;
RA   Holman G.M., Cook B.J., Nachman R.J.;
RT   "Isolation, primary structure and synthesis of leucokinins VII and
RT   VIII: the final members of this new family of cephalomyotropic
RT   peptides isolated from head extracts of Leucophaea maderae.";
RL   Comp. Biochem. Physiol. 88C:31-34(1987).
CC   -!- FUNCTION: THIS CEPHALOMYOTROPIC PEPTIDE STIMULATES CONTRACTILE
CC       ACTIVITY OF COCKROACH PROTODEUM (HINDGUT).
CC   -!- SIMILARITY: TO THE OTHER LEUCOKININS.
DR   PIR; JS0318; JS0318.
KW   Neuropeptide; Amidation.
```

```
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  902 MW;  736365AB59CAADD8 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GAN 6
           ||:
Db       1 GAD 3


RESULT  15
MOSF_CLYJA
ID   MOSF_CLYJA     STANDARD;      PRT;     9 AA.
AC   P19853;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   [PHE-6]-MOSACT.
OS   Clypeaster japonicus (Sand dollar).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Gnathostomata; Clypeasteroida;
OC   Clypeasteridae; Clypeaster.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=EGG JELLY;
RA   Suzuki N., Kurita M., Yoshino K.I., Kajiura H., Nomura K.,
RA   Yamaguchi M.;
RT   "Purification and structure of mosact and its derivatives from the
RT   egg jelly of the sea urchin Clypeaster japonicus.";
RL   Zool. Sci. 4:649-656(1987).
CC   -!- FUNCTION: STIMULATES SPERM RESPIRATION AND MOTILITY.
DR   PIR; JN0027; JN0027.
SQ   SEQUENCE   9 AA;  924 MW;  93245729CDC5BAB5 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
            :| |
Db       6 FLIG 9


RESULT  16
OXYA_SCYCA
ID   OXYA_SCYCA     STANDARD;      PRT;     9 AA.
AC   P42996;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ASVATOCIN.
OS   Scyliorhinus canicula (Spotted dogfish) (Spotted catshark).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Galeomorphii; Galeoidea; Carcharhiniformes;
OC   Scyliorhinidae; Scyliorhinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PITUITARY;
RX   MEDLINE; 95062247.
RA   Chauvet J., Rouille Y., Chauveau C., Chauvet M.-T., Acher R.;
RT   "Special evolution of neurohypophysial hormones in cartilaginous
RT   fishes: asvatocin and phasvatocin, two oxytocin-like peptides
RT   isolated from the spotted dogfish (Scyliorhinus caniculus).";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:11266-11270(1994).
CC   -!- FUNCTION: DISPLAYS OXYTOCIC ACTIVITY ON RAT UTERUS.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
```

```
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  982 MW;  17EDD76EB444404B CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   14.3%;  Pred. No. 8.8e+04;
  Matches    1;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        1 YLSGANI 7
            |::   :
Db        2 YINNCPV 8


RESULT  17
OXYT_BUFRE
ID   OXYT_BUFRE     STANDARD;      PRT;     9 AA.
AC   P42995;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   SERITOCIN.
OS   Bufo regularis (African toad).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Bufonidae;
OC   Bufo.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PITUITARY NEUROINTERMEDIATE LOBE;
RX   MEDLINE; 96059313.
RA   Chauvet J., Michel G., Ouedraogo Y., Chou J., Chait B.T., Acher R.;
RT   "A new neurohypophysial peptide, seritocin ([Ser5,Ile8]-oxytocin),
RT   identified in a dryness-resistant African toad, Bufo regularis.";
RL   Int. J. Pept. Protein Res. 45:482-487(1995).
CC   -!- FUNCTION: DEVOID OF OXYTOCIC ACTIVITY.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; FALSE_NEG.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  983 MW;  17FF476EA5A6D04B CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   28.6%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        1 YLSGANI 7
            |:    |
Db        2 YIQSCPI 8


RESULT  18
TAL3_PICJA
ID   TAL3_PICJA     STANDARD;      PRT;     9 AA.
AC   P17441;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   TRANSALDOLASE III (EC 2.2.1.2) (FRAGMENT).
OS   Pichia jadinii (Yeast) (Candida utilis).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Pichia.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 75145197.
RA   Tsolas O., Sun S.C.;
RT   "Isolation of a peptide containing a histidinyl-cysteinyl sequence
RT   from the active center of transaldolase.";
RL   Arch. Biochem. Biophys. 167:525-533(1975).
```

```
CC   -!- FUNCTION: TRANSALDOLASE IS IMPORTANT FOR THE BALANCE OF
CC       METABOLITES IN THE PENTOSE-PHOSPHATE PATHWAY.
CC   -!- CATALYTIC ACTIVITY: SEDOHEPTULOSE 7-PHOSPHATE + D-GLYCERALDEHYDE
CC       3-PHOSPHATE = D-ERYTHROSE 4-PHOSPHATE + D-FRUCTOSE 6-PHOSPHATE.
CC   -!- PATHWAY: NONOXYDATIVE PART OF THE PENTOSE-PHOSPHATE PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE TRANSALDOLASE FAMILY.
DR   PIR; A11497; A11497.
DR   INTERPRO; IPR001585; -.
DR   PROSITE; PS00958; TRANSALDOLASE_2; PARTIAL.
DR   PROSITE; PS01054; TRANSALDOLASE_1; PARTIAL.
KW   Transferase; Pentose shunt.
FT   NON_TER       1      1
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1033 MW;  325A31A44EB1E058 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   40.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        4 GANIN 8
            | : |
Db        2 GIHCN 6


RESULT  19
TPIS_CANFA
ID   TPIS_CANFA     STANDARD;      PRT;     5 AA.
AC   P54714;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   TRIOSEPHOSPHATE ISOMERASE (EC 5.3.1.1) (TIM) (FRAGMENT).
GN   TPI1.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=HEART;
RX   MEDLINE; 98163340.
RA   Dunn M.J., Corbett J.M., Wheeler C.H.;
RT   "HSC-2DPAGE and the two-dimensional gel electrophoresis database of
RT   dog heart proteins.";
RL   Electrophoresis 18:2795-2802(1997).
CC   -!- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE = DIHYDROXY-
CC       ACETONE PHOSPHATE.
CC   -!- PATHWAY: PLAYS AN IMPORTANT ROLE IN SEVERAL METABOLIC PATHWAYS.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.
DR   HSC-2DPAGE; P54714; DOG.
DR   INTERPRO; IPR000652; -.
DR   PROSITE; PS00171; TIM; PARTIAL.
KW   Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;
KW   Pentose shunt.
FT   NON_TER       1      1
FT   NON_TER       5      5
SQ   SEQUENCE   5 AA;  550 MW;  64444862C9A00000 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 5;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 GAN 6
            | |
Db        3 GNN 5


RESULT  20
UXA4_CHLTR
ID   UXA4_CHLTR     STANDARD;      PRT;     5 AA.
AC   P38005;
```

```
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE FROM ELEMENTARY BODY (FRAGMENT).
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE.
RC   STRAIN=L2/434/BU;
RA   Bini L., Santucci A., Magi B., Marzocchi B., Sanchez-Campillo M.,
RA   Comanducci M., Christianen G., Birkelund S., Vtretou E., Ratti G.,
RA   Pallini V.;
RL   Submitted (SEP-1994) to the SWISS-PROT data bank.
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 4.5, ITS MW IS: 28 KDA.
FT   NON_TER       5      5
SQ   SEQUENCE   5 AA;  474 MW;  75BAA865AA800000 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         3 SG 4
             ||
Db         2 SG 3


RESULT  21
CIP2_MYTED
ID   CIP2_MYTED     STANDARD;      PRT;     6 AA.
AC   P13737;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   CONTRACTION-INHIBITING PEPTIDE II (MIP II).
OS   Mytilus edulis (Blue mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytiloidea; Mytilidae; Mytilus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PEDAL GANGLION;
RX   MEDLINE; 88240357.
RA   Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.;
RT   "Structures and actions of Mytilus inhibitory peptides.";
RL   Biochem. Biophys. Res. Commun. 152:1376-1382(1988).
CC   -!- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN
CC       MUSCLES.
CC   -!- SIMILARITY: TO MIP I.
DR   PIR; B27696; B27696.
KW   Hormone; Amidation.
FT   MOD_RES       6      6       AMIDATION.
SQ   SEQUENCE   6 AA;  621 MW;  72C9C6876DD81000 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         4 GA 5
             ||
Db         1 GA 2


RESULT  22
TMOF_SARBU
ID   TMOF_SARBU     STANDARD;      PRT;     6 AA.
AC   P41495;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   TRYPSIN-MODULATING OOSTATIC FACTOR (TMOF).
OS   Sarcophaga bullata (Grey flesh fly) (Neobellieria bullata).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Sarcophagidae; Sarcophaga.
RN   [1]
RP   SEQUENCE, AND SYNTHESIS.
RC   TISSUE=OVARY;
RX   MEDLINE; 94211930.
RA   Bylemans D., Borovsky D., Hunt D.F., Shabanowitz J., Grauwels L.,
RA   de Loof A.;
RT   "Sequencing and characterization of trypsin modulating oostatic
RT   factor (TMOF) from the ovaries of the grey fleshfly, Neobellieria
RT   (Sarcophaga) bullata.";
RL   Regul. Pept. 50:61-72(1994).
CC   -!- FUNCTION: HAS AN OOSTATIC ACTIVITY. INHIBITS TRYPSIN BIOSYNTHESIS
CC       IN THE MIDGUT WHICH INDIRECTLY REDUCES THE VITELLOGENIN
CC       CONCENTRATION IN THE HEMOLYMPH RESULTING IN INHIBITION OF OOCYTE
CC       DEVELOPMENT.
CC   -!- DEVELOPMENTAL STAGE: SYNTHESIZED AND RELEASED FROM FOLLICULAR
CC       EPITHELIUM AFTER A BLOOD MEAL.
KW   Hormone.
SQ   SEQUENCE   6 AA;  695 MW;  61E72451B7642000 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         8 NL 9
             ||
Db         4 NL 5


RESULT  23
ALL2_CARMA
ID   ALL2_CARMA     STANDARD;      PRT;     7 AA.
AC   P81805;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 2.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   MOD_RES       7      7       AMIDATION (POTENTIAL).
SQ   SEQUENCE   7 AA;  770 MW;  672879CDCB5DDB70 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         1 YLSG 4
             |  |
Db         3 YAFG 6


RESULT  24
ALL3_CARMA
ID   ALL3_CARMA     STANDARD;      PRT;     7 AA.
AC   P81806;
```

```
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 3.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Multigene family.
SQ   SEQUENCE   7 AA;  796 MW;  672879CDCB476B70 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |  |
Db       3 YAFG 6

RESULT  25
ALL4_CARMA
ID   ALL4_CARMA     STANDARD;      PRT;     7 AA.
AC   P81807;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 4.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Multigene family.
SQ   SEQUENCE   7 AA;  782 MW;  672879CDCB476AC0 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |  |
Db       3 YAFG 6

RESULT  26
ALL5_CARMA
ID   ALL5_CARMA     STANDARD;      PRT;     7 AA.
AC   P81808;
DT   30-MAY-2000 (Rel. 39, Created)
```

```
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 5.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   MOD_RES       7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  781 MW;  672879CDCB476420 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |  |
Db       3 YAFG 6

RESULT  27
UN06_PINPS
ID   UN06_PINPS     STANDARD;      PRT;     7 AA.
AC   P81675;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE OF NEEDLES (N141) (FRAGMENT).
OS   Pinus pinaster (Maritime pine).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Coniferopsida; Coniferales; Pinaceae; Pinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=NEEDLE;
RX   MEDLINE; 99274088.
RA   Costa P., Pionneau C., Bauw G., Dubos C., Bahrman N., Kremer A.,
RA   Frigerio J.-M., Plomion C.;
RT   "Separation and characterization of needle and xylem maritime pine
RT   proteins.";
RL   Electrophoresis 20:1098-1108(1999).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 6.6, ITS MW IS: 25 KDA.
FT   NON_TER       1      1
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  823 MW;  69D76724486B5740 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 NL 9
           ||
Db       4 NL 5

RESULT  28
AL12_CARMA
ID   AL12_CARMA     STANDARD;      PRT;     8 AA.
AC   P81815;
DT   30-MAY-2000 (Rel. 39, Created)
```

```
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 12.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Multigene family.
SQ   SEQUENCE   8 AA;  913 MW;  672879CDCB569AB7 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |  |
Db       4 YAFG 7


RESULT  29
AL17_CARMA
ID   AL17_CARMA     STANDARD;      PRT;     8 AA.
AC   P81820;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 17.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   MOD_RES       8      8       AMIDATION (POTENTIAL).
SQ   SEQUENCE   8 AA;  858 MW;  C82879D5AB46D865 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SG 4
           ||
Db       1 SG 2


RESULT  30
ALL1_CYDPO
ID   ALL1_CYDPO     STANDARD;      PRT;     8 AA.
AC   P82152;
DT   30-MAY-2000 (Rel. 39, Created)
```

```
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYDIASTATIN 1.
OS   Cydia pomonella (Codling moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Tortricoidea; Tortricidae; Olethreutinae; Cydia.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LARVA;
RX   MEDLINE; 98054539.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
RA   Davey M., East P.D., Thorpe A.;
RT   "Lepidopteran peptides of the allatostatin superfamily.";
RL   Peptides 18:1301-1309(1997).
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  934 MW;  C82879C45B51F775 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 NINL 9
             |  |
Db       5 NFGL 8


RESULT  31
ALL6_CYDPO
ID   ALL6_CYDPO     STANDARD;      PRT;     8 AA.
AC   P82157;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYDIASTATIN 6.
OS   Cydia pomonella (Codling moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Tortricoidea; Tortricidae; Olethreutinae; Cydia.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LARVA;
RX   MEDLINE; 98054539.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
RA   Davey M., East P.D., Thorpe A.;
RT   "Lepidopteran peptides of the allatostatin superfamily.";
RL   Peptides 18:1301-1309(1997).
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  936 MW;  0B2879C45B573767 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 NINL 9
             |  |
Db       5 NFGL 8


RESULT  32
ALL7_CARMA
ID   ALL7_CARMA     STANDARD;      PRT;     8 AA.
AC   P81809; P81810; P81804;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 7 [CONTAINS: CARCINUSTATIN 6; CARCINUSTATIN 1].
```

```
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   CHAIN         1      8       CARCINUSTATIN 7.
FT   CHAIN         2      8       CARCINUSTATIN 6.
FT   CHAIN         4      8       CARCINUSTATIN 1.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  825 MW;  922879CDCB4775BD CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |  |
Db       4 YAFG 7


RESULT  33
ALL8_CARMA
ID   ALL8_CARMA     STANDARD;      PRT;     8 AA.
AC   P81811;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 8.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  795 MW;  922879CDCB47687D CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |  |
Db       4 YAFG 7


RESULT  34
ALL9_CARMA
ID   ALL9_CARMA     STANDARD;      PRT;     8 AA.
AC   P81812;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 9.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  781 MW;  7C2879CDCB476878 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |  |
Db       4 YAFG 7


RESULT  35
B44K_PORGI
ID   B44K_PORGI     STANDARD;      PRT;     8 AA.
AC   P81886;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   44 KDA IMMUNOGENIC PROTEIN (FRAGMENT).
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; CFB group; Bacteroidaceae; Porphyromonas.
RN   [1]
RP   SEQUENCE.
RC   STRAIN=VPB 3492;
RX   MEDLINE; 20198497.
RA   Norris J.M., Love D.N.;
RT   "Serum antibody responses of cats to soluble whole cell antigens of
RT   feline porphyromonas gingivalis.";
RL   Vet. Microbiol. 73:37-49(2000).
CC   -!- SIMILARITY: TO P.GINGIVALIS HEMAGGLUTININ A.
KW   Antigen.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  989 MW;  9554540326CB476D CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 NI 7
           ||
Db       7 NI 8


Search completed: December 16, 2000, 04:23:31
Job time: 4566 sec
```

```
                     GenCore version 4.5
            Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 03:35:12 ; Search time 89.11 Seconds
                                              (without alignments)
                                              6.409 Million cell updates/sec

Title:           US-09-529-121-4
Perfect score:   45
Sequence:        1 YLSGANINL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters:      793

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       PIR_65:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                      %
Result              Query
  No.    Score      Match Length DB  ID                       Description
-------------------------------------------------------------------------------
     1       16      35.6      8  2  H41978                   callFMRFamide 8 -
     2       16      35.6      8  2  E60588                   sperm-activating p
     3       15      33.3      8  2  T13818                   cytochrome oxidase
     4       15      33.3      9  2  A61364                   isotocin - common
     5       15      33.3      9  2  PT0288                   Ig heavy chain CRD
     6       15      33.3      9  2  G41946                   T-cell receptor ga
     7       14      31.1      5  2  S62883                   seminal plasma pro
     8       14      31.1      6  2  PT0605                   T-cell receptor be
     9       14      31.1      6  2  PT0593                   T-cell receptor be
    10       14      31.1      7  2  A34818                   vicilin 72K chain
    11       14      31.1      7  2  PT0654                   T-cell receptor be
    12       14      31.1      7  2  S58797                   serine/threonine-s
    13       14      31.1      8  2  A21440                   variant surface gl
    14       14      31.1      8  2  PN0043                   phosphatidylethano
    15       14      31.1      9  2  C57444                   neuropeptide Grb-A
    16       14      31.1      9  2  JQ0914                   MHC class I histoc
    17       14      31.1      9  2  S56004                   glucan 1,3-beta-gl
    18       13      28.9      4  2  S43959                   Ig mu chain V regi
    19       13      28.9      7  2  I40504                   hypothetical prote
    20       13      28.9      7  2  I50210                   gene c-rel protein
    21       13      28.9      9  1  YFPG                     thymic factor - pi
    22       13      28.9      9  2  A60957                   thymocyte growth p
    23       13      28.9      9  2  C41170                   photosystem II pro
    24       13      28.9      9  2  PT0268                   Ig heavy chain CRD
    25       12      26.7      6  2  I51434                   H4 histone - Afric
    26       12      26.7      7  1  XEYDGD                   galactose oxidase
    27       12      26.7      8  2  PQ0701                   unidentified 6.5/3
    28       12      26.7      8  2  PL0184                   capsid protein VP-
    29       12      26.7      9  2  A44873                   caldesmon - rabbit
    30       11      24.4      4  2  S43014                   hypothetical prote
    31       11      24.4      5  1  HOROHA                   proctolin - Americ
    32       11      24.4      5  2  A41225                   copper resistance
    33       11      24.4      5  2  A60411                   proctolin - Atlant
    34       11      24.4      5  2  C23751                   spinal cord peptid
    35       11      24.4      5  2  G44817                   27.5 kda structura
    36       11      24.4      5  2  I44817                   27.5K structural p
    37       11      24.4      5  2  E44817                   27.5K structural p
    38       11      24.4      5  2  C44817                   28.5K structural p
    39       11      24.4      5  2  A44817                   28K structural pro
    40       11      24.4      6  2  JU0355                   lipopeptide WS1279
    41       11      24.4      6  2  B44510                   hypothetical prote
    42       11      24.4      6  2  PT0280                   Ig heavy chain CRD
    43       11      24.4      6  2  I49424                   cytotoxic T-lympho
    44       11      24.4      7  2  PQ0728                   unidentified 5.0/1
    45       11      24.4      7  2  S45311                   microcin C7 - Esch
    46       11      24.4      7  2  S78024                   ribosomal protein
    47       11      24.4      7  2  H33098                   180K exoantigen -
    48       11      24.4      7  2  PT0542                   T-cell receptor be
    49       11      24.4      7  2  A58718                   carnocin UI49 - Ca
    50       11      24.4      8  2  JS0318                   leucokinin VIII -
    51       11      24.4      8  2  A41117                   acetylcholinestera
    52       11      24.4      8  2  A61467                   penalbumin - Adeli
    53       11      24.4      8  2  PT0311                   Ig heavy chain CRD
    54       11      24.4      8  2  A42689                   major postsynaptic
    55       11      24.4      8  2  B47594                   aspartate kinase (
    56       11      24.4      8  2  S68325                   blood cell protein
    57       11      24.4      9  2  A61230                   calsequestrin, car
    58       11      24.4      9  2  S63491                   dissimilatory sulf
    59       11      24.4      9  2  A43848                   cell surface adhes
    60       11      24.4      9  2  A11497                   transaldolase (EC
    61       11      24.4      9  2  A57444                   neuropeptide Grb-A
    62       11      24.4      9  2  B57444                   neuropeptide Grb-A
    63       11      24.4      9  2  JN0027                   [Phe-6]-mosact - s
    64       11      24.4      9  2  I50633                   c-rel protein - ch
    65       11      24.4      9  2  A60427                   macrophage cytotox
    66       11      24.4      9  2  PH0935                   T-cell receptor be
    67       11      24.4      9  2  PH0918                   T-cell receptor be
    68       11      24.4      9  4  I57650                   hemoglobin alpha c
    69       10      22.2      3  2  PT0571                   T-cell receptor be
    70       10      22.2      4  2  PT0271                   Ig heavy chain CRD
    71       10      22.2      4  2  A53284                   T-cell receptor be
    72       10      22.2      4  2  PT0633                   T-cell receptor be
    73       10      22.2      4  2  PT0711                   T-cell receptor be
    74       10      22.2      4  2  PT0698                   T-cell receptor be
    75       10      22.2      4  2  PT0677                   T-cell receptor be


                                ALIGNMENTS


RESULT   1
H41978
calliFMRFamide 8 - bluebottle fly (Calliphora vomitoria)
C;Species: Calliphora vomitoria
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C;Accession: H41978
R;Duve, H.; Johnsen, A.H.; Sewell, J.C.; Scott, A.G.; Orchard, I.; Rehfeld, J.F.; Tho
Proc. Natl. Acad. Sci. U.S.A. 89, 2326-2330, 1992
A;Title: Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH-2 neuropeptides (d
A;Reference number: A41978; MUID:92196111
A;Accession: H41978
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-8 <DUV>
C;Keywords: amidated carboxyl end; neuropeptide
F;8/Modified site: amidated carboxyl end (Phe) #status experimental


  Query Match            35.6%;  Score 16;  DB 2;  Length 8;
  Best Local Similarity  100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       4 GAN 6
           |||
Db       1 GAN 3


RESULT   2
E60588
sperm-activating peptide a - sea urchin (Pseudoboletia maculata)
N;Alternate names: speract homolog
C;Species: Pseudoboletia maculata
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 20-Jun-2000
C;Accession: E60588
R;Yoshino, K.I.; Kajiura, H.; Nomura, K.; Takao, T.; Shimonishi, Y.; Kurita, M.; Yamaguc
Comp. Biochem. Physiol. B 94, 739-751, 1989
A;Title: A halogenated amino acid-containing sperm activating peptide and its related pe
otus nudus, Echinometra mathaei and Heterocentrotus mammillatus.
A;Reference number: A60527
A;Accession: E60588
A;Molecule type: protein
A;Residues: 1-8 <YOS>


  Query Match             35.6%;  Score 16;  DB 2;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LSGAN 6
           | | |
Db       4 LDGVN 8


RESULT   3
T13818
cytochrome oxidase subunit I - Atlantic hagfish mitochondrion (fragment)
C;Species: mitochondrion Myxine glutinosa (Atlantic hagfish)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C;Accession: T13818
R;Delarbre, C.; Barriel, V.; Tillier, S.; Janvier, P.; Gachelin, G.
Mol. Biol. Evol. 14, 807-813, 1997
A;Title: The main features of the craniate mitochondrial DNA between the ND1 and the COI
A;Reference number: Z17775; MUID:97398704
A;Accession: T13818
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-8 <DEL>
A;Cross-references: EMBL:Y09527; NID:e1011465; PID:e329906; PIDN:CAA70718.1
C;Genetics:
A;Genome: mitochondrion
A;Note: COI
C;Keywords: mitochondrion


  Query Match             33.3%;  Score 15;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLS 3
           |||
Db       2 YLS 4


RESULT   4
A61364
isotocin - common carp
C;Species: Cyprinus carpio (common carp)
C;Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 20-Jun-2000
C;Accession: A61364
R;Acher, R.; Chauvet, J.; Chauvet, M.T.; Crepy, D.
Comp. Biochem. Physiol. A 14, 245-254, 1965
A;Title: Caracterisation des hormones neurohypophysaires d'un poisson osseux d'eau douce
A;Reference number: A61364
A;Accession: A61364
A;Status: preliminary
```

```
A;Molecule type: protein
A;Residues: 1-9 <ACH>
C;Superfamily: oxytocin-neurophysin
C;Keywords: amidated carboxyl end; neuropeptide; posterior pituitary
F;9/Modified site: amidated carboxyl end (Gly) #status experimental


  Query Match             33.3%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity   42.9%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 YLSGANI 7
           |:|   |
Db       2 YISNCPI 8


RESULT   5
PT0288
Ig heavy chain CRD3 region (clone 4-106) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C;Accession: PT0288
R;Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991
A;Title: Preferential utilization of specific immunoglobulin heavy chain diversity an
A;Reference number: PT0222; MUID:91108337
A;Accession: PT0288
A;Molecule type: DNA
A;Residues: 1-9 <YAM>
A;Experimental source: B lymphocyte
C;Keywords: heterotetramer; immunoglobulin


  Query Match             33.3%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           | ||
Db       5 YSSG 8


RESULT   6
G41946
T-cell receptor gamma chain (2t.23) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 07-May-1999
C;Accession: G41946
R;Whetsell, M.; Mosley, R.L.; Whetsell, L.; Schaefer, F.V.; Miller, K.S.; Klein, J.R.
Mol. Cell. Biol. 11, 5902-5909, 1991
A;Title: Rearrangement and junctional-site sequence analyses of T-cell receptor gamma
A;Reference number: A41946; MUID:92049316
A;Accession: G41946
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-9 <WHE>
C;Keywords: T-cell receptor


  Query Match             33.3%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           | ||
Db       5 YSSG 8


RESULT   7
S62883
seminal plasma protein II - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
```

```
C;Accession: S62883
R;Romero, A.; Varela, P.F.; Sanz, L.; Toepfer-Petersen, E.; Calvete, J.J.
FEBS Lett. 382, 15-17, 1996
A;Title: Crystallization and preliminary X-ray diffraction analysis of boar seminal plas
A;Reference number: S62882; MUID:96196555
A;Accession: S62883
A;Molecule type: protein
A;Residues: 1-5 <ROM>
C;Complex: heterodimer; seminal plasma protein I and seminal plasma protein II
C;Keywords: glycoprotein; heterodimer; semen


  Query Match             31.1%;  Score 14;  DB 2;  Length 5;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 ANIN 8
           | ||
Db       1 ARIN 4

RESULT   8
PT0605
T-cell receptor beta chain V-D-J region (120-1L) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0605
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0605
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-6 <FEE>
A;Experimental source: newborn thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGA 5
           |||
Db       2 SGA 4

RESULT   9
PT0593
T-cell receptor beta chain V-D-J region (159-1F) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0593
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0593
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-6 <FEE>
A;Experimental source: day 19 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGA 5
           |||
Db       4 SGA 6
```

```
RESULT  10
A34818
vicilin 72K chain - pigeon pea (fragment)
C;Species: Cajanus cajan (pigeon pea)
C;Date: 13-Jul-1990 #sequence_revision 13-Jul-1990 #text_change 30-Sep-1993
C;Accession: A34818
R;Mawal, Y.R.; Mawal, M.R.; Ranjekar, P.K.
Biochem. Biophys. Res. Commun. 166, 1446-1452, 1990
A;Title: Unusual denaturation properties of vicilin from Cajanus cajan.
A;Reference number: A34818; MUID:90165956
A;Accession: A34818
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-7 <MAW>


  Query Match             31.1%;  Score 14;  DB 2;  Length 7;
  Best Local Similarity   40.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GANIN 8
           || ::
Db       1 GARVD 5

RESULT  11
PT0654
T-cell receptor beta chain V-D-J region (121-1BK) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0654
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0654
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-7 <FEE>
A;Experimental source: day 4 postnatal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGA 5
           |||
Db       2 SGA 4

RESULT  12
S58797
serine/threonine-specific protein kinase c-mos - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 23-Feb-1997
C;Accession: S58797
R;Nagao, Y.
Biochim. Biophys. Acta 1245, 130-143, 1995
A;Title: Expression of c-mos protein in cultured rat spermatogenic cells and evidence
A;Reference number: S58797; MUID:95383384
A;Accession: S58797
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-7 <NAG>
C;Genetics:
A;Gene: c-mos
C;Keywords: phosphotransferase; protein kinase
```

```
  Query Match             31.1%;  Score 14;  DB 2;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 GANI 7
              | |:
Db          3 GGNL 6

RESULT  13
A21440
variant surface glycoprotein pSLc1 - Trypanosoma brucei (fragment)
C;Species: Trypanosoma brucei
C;Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 20-Mar-1998
C;Accession: A21440
R;Parsons, M.; Nelson, R.G.; Watkins, K.P.; Agabian, N.
Cell 38, 309-316, 1984
A;Title: Trypanosome mRNAs share a common 5' spliced leader sequence.
A;Reference number: A90853; MUID:84282716
A;Accession: A21440
A;Molecule type: mRNA
A;Residues: 1-8 <PAR>
A;Cross-references: GB:K02195; NID:g162150; PID:g162151
C;Keywords: glycoprotein

  Query Match             31.1%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   33.3%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 LSGANI 7
              :||  :
Db          1 MSGKEV 6

RESULT  14
PN0043
phosphatidylethanol amine-binding protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 29-Oct-1997 #sequence_revision 29-Oct-1997 #text_change 23-Jan-1998
C;Accession: PN0043
R;Kato, H.
Kawasaki Igakkaishi 22, 245-259, 1996
A;Title: Analysis of proteins isolated by two dimensional electrophoresis of mouse neur
A;Reference number: PN0041
A;Accession: PN0043
A;Molecule type: protein
A;Residues: 1-8 <KAT>
A;Experimental source: neuroblastoma cell
C;Comment: The molecular mass is 23,500 and the pI is 4.8. The amino-terminus is blocke
C;Keywords: brain

  Query Match             31.1%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LSG 4
              |||
Db          5 LSG 7

RESULT  15
C57444
neuropeptide Grb-AST B3 - two-spotted cricket
C;Species: Gryllus bimaculatus (two-spotted cricket)
C;Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 26-Jan-1996
C;Accession: C57444
R;Lorenz, M.W.; Kellner, R.; Hoffmann, K.H.
J. Biol. Chem. 270, 21103-21108, 1995
A;Title: A family of neuropeptides that inhibit juvenile hormone biosynthesis in the cri
A;Reference number: A57444; MUID:95403341
A;Accession: C57444
```

```
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <LOR>

  Query Match             31.1%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LSG 4
              |||
Db          5 LSG 7

RESULT  16
JQ0914
MHC class I histocompatibility antigen heavy chain - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 24-Feb-1995
C;Accession: JQ0914
R;Kanki, T.; Kuwasawa, N.; Sekiya, Y.; Ichikawa, Y.
submitted to JIPID, May 1991
A;Description: Responsive expression of a MHC class I epitope and genes following Mar
A;Reference number: JQ0914
A;Accession: JQ0914
A;Molecule type: mRNA
A;Residues: 1-9 <KAN>
A;Experimental source: kidney, strain cornell N

  Query Match             31.1%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 SGAN 6
              :|:|
Db          3 TGSN 6

RESULT  17
S56004
glucan 1,3-beta-glucosidase (EC 3.2.1.58) GNIII, extracellular - fungus (Acremonium p
N;Alternate names: (1-3)-beta-D-glucan glucohydrolase GNIII
C;Species: Acremonium persicinum
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 07-May-1999
C;Accession: S56004
R;Pitson, S.M.; Seviour, R.J.; McDougall, B.M.; Woodward, J.R.; Stone, B.A.
Biochem. J. 308, 733-741, 1995
A;Title: Purification and characterization of three extracellular (1->3)-beta-D-gluca
A;Reference number: S56002; MUID:97104268
A;Accession: S56004
A;Molecule type: protein
A;Residues: 1-9 <PIT>
C;Keywords: glycosidase; hydrolase

  Query Match             31.1%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 ANI 7
              |||
Db          7 ANI 9

RESULT  18
S43959
Ig mu chain V region (clone 13) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 20-Oct-1994 #sequence_revision 26-Jul-1996 #text_change 17-Mar-1999
C;Accession: S43959
R;Wagner, S.D.; Williams, G.T.; Larson, T.; Neuberger, M.S.; Kitamura, D.; Rajewsky,
Nucleic Acids Res. 22, 1389-1393, 1994
```

```
A;Title: Antibodies generated from human immunoglobulin miniloci in transgenic mice.
A;Reference number: S43956; MUID:94248036
A;Accession: S43959
A;Molecule type: DNA
A;Residues: 1-4 <WAG>
C;Keywords: immunoglobulin


  Query Match             28.9%;  Score 13;  DB 2;  Length 4;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSG 4
            | :|
Db        1 YCAG 4


RESULT  19
I40504
hypothetical protein 2 (7 aa) - Bacillus stearothermophilus
C;Species: Bacillus stearothermophilus
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Oct-1999
C;Accession: I40504
R;Waye, M.M.; Winter, G.
Eur. J. Biochem. 158, 505-510, 1986
A;Title: A transcription terminator in the 5' non-coding region of the tyrosyl tRNA synt
A;Reference number: I40503; MUID:86274732
A;Accession: I40504
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-7 <RES>
A;Cross-references: EMBL:X04193; NID:g40233; PIDN:CAA27782.1; PID:g580943


  Query Match             28.9%;  Score 13;  DB 2;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 ANI 7
            ||:
Db        2 ANV 4


RESULT  20
I50210
gene c-rel protein - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 28-Feb-1997
C;Accession: I50210
R;Kabrun, N.; Bumstead, N.; Hayman, M.J.; Enrietto, P.J.
Mol. Cell. Biol. 10, 4788-4794, 1990
A;Title: Characterization of a novel promoter insertion in the c-rel locus.
A;Reference number: I50210; MUID:90355995
A;Accession: I50210
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-7 <KAB>
A;Cross-references: GB:M55577; NID:g555438; PID:g211661
C;Genetics:
A;Gene: c-rel


  Query Match             28.9%;  Score 13;  DB 2;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSGA 5
            ::||
Db        1 MAGA 4


RESULT  21
YFPG
```

```
thymic factor - pig
N;Alternate names: FTS (facteur thymique serique)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 07-May-1999
C;Accession: A01523; A60983
R;Pleau, J.M.; Dardenne, M.; Blouquit, Y.; Bach, J.F.
J. Biol. Chem. 252, 8045-8047, 1977
A;Title: Structural study of circulating thymic factor: a peptide isolated from pig s
A;Reference number: A01523; MUID:78026571
A;Accession: A01523
A;Molecule type: protein
A;Residues: 1-9 <PLE>
R;Bach, J.F.; Dardenne, M.; Pleau, J.M.; Rosa, J.
Nature 266, 55-57, 1977
A;Title: Biochemical characterisation of a serum thymic factor.
A;Reference number: A60983; MUID:77123829
A;Accession: A60983
A;Molecule type: protein
A;Residues: 'Z',2-4,'Z',6-9 <BAC>
C;Comment: This peptide induces DNA synthesis in immature thymocytes, but not periphe
in a variety of immunoassays.
C;Comment: See PIR:A60957 (sheep) for discussion of another possible N-terminal modif
C;Superfamily: thymic factor
C;Keywords: pyroglutamic acid
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental


  Query Match             28.9%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        4 GAN 6
            |:|
Db        7 GSN 9


RESULT  22
A60957
thymocyte growth peptide - sheep
N;Contains: FTS (facteur thymique serique)
C;Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 24-Nov-1999
C;Accession: A60957
R;Ernstroem, U.; Gafvelin, G.; Rudja, J.M.
Biosci. Rep. 10, 403-412, 1990
A;Title: Purification of thymocyte growth peptide (TGP) from sheep thymus. Relationsh
A;Reference number: A60957; MUID:91064427
A;Accession: A60957
A;Molecule type: protein
A;Residues: 1-9 <ERN>
C;Comment: This peptide induces DNA synthesis in immature thymocytes, but not periphe
in a variety of immunoassays.
C;Comment: This peptide was isolated in two forms. One form contained the pyrrolidone
r form (thymocyte growth peptide) contains a large, non-peptide blocking group with a
C;Superfamily: thymic factor
C;Keywords: blocked amino end; pyroglutamic acid
F;1/Modified site: pyrrolidone carboxylic acid (Glx) (in FTS) #status experimental
F;1/Modified site: blocked amino end (Glx) (in thymocyte growth peptide) #status expe


  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        4 GAN 6
            |:|
Db        7 GSN 9


RESULT  23
C41170
photosystem II protein psbM - Chlamydomonas reinhardtii (fragment)
C;Species: Chlamydomonas reinhardtii
```

```
C;Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 30-Sep-1993
C;Accession: C41170
R;de Vitry, C.; Diner, B.A.; Popot, J.L.
J. Biol. Chem. 266, 16614-16621, 1991
A;Title: Photosystem II particles from Chlamydomonas reinhardtii. Purification, molecula
A;Reference number: A41170; MUID:91358452
A;Accession: C41170
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <DE5>


  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGA 5
           ::||
Db       4 IAGA 7


RESULT  24
PT0268
Ig heavy chain CRD3 region (clone 3-94B) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C;Accession: PT0268
R;Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991
A;Title: Preferential utilization of specific immunoglobulin heavy chain diversity and j
A;Reference number: PT0222; MUID:91108337
A;Accession: PT0268
A;Molecule type: DNA
A;Residues: 1-9 <YAM>
A;Experimental source: B lymphocyte
C;Keywords: heterotetramer; immunoglobulin


  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   60.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 GANIN 8
           |  ||
Db       4 GIPIN 8


RESULT  25
I51434
H4 histone - African clawed frog (fragment)
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Sep-1996
C;Accession: I51434
R;Woodland, H.R.; Warmington, J.R.; Ballantine, J.E.M.; Turner, P.C.
Nucleic Acids Res. 12, 4939-4958, 1984
A;Title: Are there major developmentally regulated H4 gene classes in Xenopus?.
A;Reference number: I51391; MUID:84247348
A;Accession: I51434
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-6 <WOO>
A;Cross-references: GB:K02304; NID:g214227; PID:g555517


  Query Match             26.7%;  Score 12;  DB 2;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       1 MSG 3
```

```
RESULT  26
XEYDGD
galactose oxidase inhibitor - fungus (Cladobotryum dendroides)
C;Species: Cladobotryum dendroides
C;Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 31-Dec-1993
C;Accession: A01341
R;Avigad, G.; Markus, Z.
Fed. Proc. 31, 447, 1972
A;Reference number: A01341
A;Accession: A01341
A;Molecule type: protein
A;Residues: 1-7 <AVI>
C;Comment: The mycelia of this imperfect fungus produce the metalloenzyme galactose o
apoenzyme, may inactivate the enzyme by binding to its prosthetic copper group.
C;Superfamily: galactose oxidase inhibitor
C;Keywords: copper


  Query Match             26.7%;  Score 12;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAN 6
            :| |
Db       1 AGQN 4


RESULT  27
PQ0701
unidentified 6.5/31K protein [imported] - rice (fragment)
C;Species: Oryza sativa (rice)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: PQ0701
R;Komatsu, S.; Kajiwara, H.; Hirano, H.
Theor. Appl. Genet. 86, 935-942, 1993
A;Title: A rice protein library; a data-file of rice proteins separated by two-dimens
A;Reference number: PQ0696
A;Accession: PQ0701
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-8 <KOM>


  Query Match             26.7%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGA 5
           ::||
Db       5 VTGA 8


RESULT  28
PL0184
capsid protein VP-1 - murine poliovirus (fragment)
C;Species: murine poliovirus, Theiler's encephalomyelitis virus
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Feb-1995
C;Accession: PL0184
R;Zurbriggen, A.; Hogle, J.M.; Fujinami, R.S.
J. Exp. Med. 170, 2037-2049, 1989
A;Title: Alteration of amino acid 101 within capsid protein VP-1 changes the pathogen
A;Reference number: PL0184; MUID:90063468
A;Accession: PL0184
A;Molecule type: genomic RNA
A;Residues: 1-8 <ZUR>
C;Keywords: capsid protein


  Query Match             26.7%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 SGANIN 8
```

```
          ||   |
Db      1 SGGITN 6

RESULT  29
A44873
caldesmon - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 31-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 31-Oct-1997
C;Accession: A44873
R;Ikebe, M.; Hornick, T.
Arch. Biochem. Biophys. 288, 538-542, 1991
A;Title: Determination of the phosphorylation sites of smooth muscle caldesmon by protei
A;Reference number: A44873; MUID:91378498
A;Accession: A44873
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <IKE>
A;Experimental source: skeletal myosin
A;Note: sequence extracted from NCBI backbone (NCBIP:63199)
C;Superfamily: caldesmon


  Query Match             26.7%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   16.7%;  Pred. No. 1.8e+05;
  Matches    1;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 GANINL 9
          |::: :
Db      1 GSSLKI 6


RESULT  30
S43014
hypothetical protein URF-2Y - Yersinia enterocolitica transposon TN3926
C;Species: Yersinia enterocolitica
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
C;Accession: S43014
R;Osbourn, S.E.V.; Turner, A.K.; Grinsted, J.
submitted to the EMBL Data Library, March 1994
A;Description: The structure of the bacterial transposable element, Tn3926.
A;Reference number: S43011
A;Accession: S43014
A;Molecule type: DNA
A;Residues: 1-4 <OSB>
A;Cross-references: EMBL:X78059; NID:g460067; PIDN:CAA54980.1; PID:g581836
C;Genetics:
A;Mobile element: transposon TN3926


  Query Match             24.4%;  Score 11;  DB 2;  Length 4;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      6 NIN 8
          | |
Db      2 NAN 4


RESULT  31
HOROHA
proctolin - American cockroach
C;Species: Periplaneta americana (American cockroach)
C;Date: 29-Jul-1981 #sequence_revision 29-Jul-1981 #text_change 23-Aug-1996
C;Accession: A01644
R;Starratt, A.N.; Brown, B.E.
Life Sci. 17, 1253-1256, 1975
A;Title: Structure of the pentapeptide proctolin, a proposed neurotransmitter in insects
A;Reference number: A93048; MUID:76074708
A;Accession: A01644
A;Molecule type: protein
A;Residues: 1-5 <STA>
A;Note: the synthetic peptide had the same chromatographic, electrophoretic, and pharmac
```

```
R;O'Shea, M.; Adams, M.E.
Science 213, 567-569, 1981
A;Title: Pentapeptide (proctolin) associated with an identified neuron.
A;Reference number: A94260; MUID:81225865
A;Contents: annotation; biological source
C;Comment: This peptide is found in the lateral white neurons, which occur (in the co
innervate the striated hindgut muscles in insects and stimulate contraction of these
C;Superfamily: proctolin
C;Keywords: neuropeptide


  Query Match             24.4%;  Score 11;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YL 2
          ||
Db      2 YL 3


RESULT  32
A41225
copper resistance protein A - Pseudomonas syringae pv. tomato (fragment)
C;Species: Pseudomonas syringae pv. tomato
C;Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 24-Jun-1993
C;Accession: A41225
R;Cha, J.S.; Cooksey, D.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 8915-8919, 1991
A;Title: Copper resistance in Pseudomonas syringae mediated by periplasmic and outer
A;Reference number: A41225; MUID:92020961
A;Accession: A41225
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-5 <CHA>


  Query Match             24.4%;  Score 11;  DB 2;  Length 5;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 SGA 5
          ||:
Db      2 SGS 4


RESULT  33
A60411
proctolin - Atlantic horseshoe crab
C;Species: Limulus polyphemus (Atlantic horseshoe crab)
C;Date: 03-Feb-1993 #sequence_revision 03-Feb-1993 #text_change 17-Mar-1999
C;Accession: A60411
R;Groome, J.R.; Tillinghast, E.K.; Townley, M.A.; Vetrovs, A.; Watson III, W.H.; Hunt
Peptides 11, 205-211, 1990
A;Title: Identification of proctolin in the central nervous system of the horseshoe c
A;Reference number: A60411; MUID:90287800
A;Accession: A60411
A;Molecule type: protein
A;Residues: 1-5 <GRO>
C;Comment: This neuropeptide stimulates cardiac output and hindgut motility in the ho
C;Keywords: neuropeptide


  Query Match             24.4%;  Score 11;  DB 2;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YL 2
          ||
Db      2 YL 3


RESULT  34
C23751
```

```
spinal cord peptide SCP-6 - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 28-Sep-1987 #sequence_revision 28-Sep-1987 #text_change 30-Sep-1993
C;Accession: C23751
R;Hsi, K.L.; Chen, R.L.; Chen, Z.G.; Zhang, H.L.; Lu, Y.A.; Guo, S.Y.; Wu, S.X.; Tsou, K
Arch. Biochem. Biophys. 240, 178-183, 1985
A;Reference number: A23751; MUID:85250425
A;Accession: C23751
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-5 <HSI>


  Query Match             24.4%;  Score 11;  DB 2;  Length 5;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 SGA 5
            :||
Db        2 AGA 4


RESULT  35
G44817
27.5 kda structural protein - Leuconostoc oenos phage P32 (fragment)
C;Species: Leuconostoc oenos phage P32
C;Date: 31-Mar-1993 #sequence_revision 22-May-1998 #text_change 22-May-1998
C;Accession: G44817
R;Arendt, E.K.; Lonvaud, A.; Hammes, W.P.
J. Gen. Microbiol. 137, 2135-2139, 1991
A;Title: Lysogeny in Leuconostoc oenos.
A;Reference number: A44817; MUID:92085033
A;Accession: G44817
A;Molecule type: protein
A;Residues: 1-5 <ARE>
A;Note: sequence extracted from NCBI backbone (NCBIP:70333)


  Query Match             24.4%;  Score 11;  DB 2;  Length 5;
  Best Local Similarity   40.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        5 ANINL 9
            | : |
Db        1 ATVGL 5


Search completed: December 16, 2000, 03:35:12
Job time: 5645 sec
```

```
                        GenCore version 4.5
                Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 03:07:18 ; Search time 107.12 Seconds
                                              (without alignments)
                                              2.873 Million cell updates/sec

Title:          US-09-529-121-4
Perfect score:  45
Sequence:       1 YLSGANINL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters:      61695

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :      A_Geneseq_36:*
               1:  /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
               2:  /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
               3:  /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
               4:  /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
               5:  /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
               6:  /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
               7:  /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
               8:  /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
               9:  /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
              10:  /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
              11:  /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
              12:  /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
              13:  /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
              14:  /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
              15:  /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
              16:  /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
              17:  /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
              18:  /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
              19:  /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
              20:  /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
              21:  /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.

                              SUMMARIES
              %
Result      Query
  No. Score Match Length DB ID                     Description
-------------------------------------------------------------------------
   1    45  100.0     9  20  Y09528                Carcinoembryonic a
   2    43   95.6     9  18  W39723                Human carcina-embr
   3    43   95.6     9  19  W77134                CEA synthetic pept
   4    43   95.6     9  19  W70045                CEA derived HLA-A2
   5    43   95.6     9  20  Y47655                Immunogenic peptid
   6    43   95.6     9  20  Y09525                Carcinoembryonic a
   7    40   88.9     9  20  Y09527                Carcinoembryonic a
   8    40   88.9     9  21  Y54173                HLA binding peptid
   9    38   84.4     9  20  Y09526                Carcinoembryonic a
  10    36   80.0     9  17  W00680                Peptide comprising
  11    34   75.6     9  20  Y09529                Carcinoembryonic a
  12    27   60.0     9  17  W00690                NCA analogue of re
  13    23   51.1     9  19  W54298                Human cytohesin-1
  14    22   48.9     9  18  W19861                Fragment of enzyme
  15    21   46.7     6  11  R09414                LFA-1 alpha subuni
  16    21   46.7     6  17  R88476                Internal tryptic p
  17    21   46.7     8  13  R29031                Peptide encoded by
  18    21   46.7     8  16  R78208                B. thuringiensis e
  19    21   46.7     9  15  R73808                Antigen fragment 1
  20    21   46.7     9  17  W49339                Human leucocyte an
  21    21   46.7     9  17  W49340                Human leucocyte an
  22    20   44.4     6  21  Y83831                Pyrrolobenzodiazep
  23    20   44.4     7  20  Y41846                Rheumatoid arthrit
  24    20   44.4     7  21  Y54382                Amino acid sequenc
  25    20   44.4     9  16  R69993                Nonameric mimotope
  26    20   44.4     9  19  W54299                Human cytohesin-1
  27    20   44.4     9  19  W40267                K. oxytoca R-speci
  28    19   42.2     6  20  Y23411                V beta 6 clone fou
  29    19   42.2     7  16  R75676                Streptokinase-bind
  30    19   42.2     7  18  W12026                Herpes virus DNA p
  31    19   42.2     8  14  R34629                B. thuringiensis d
  32    19   42.2     8  16  R78218                B. thuringiensis e
  33    19   42.2     9  13  R23738                Ligand CDR(f).  Sy
  34    19   42.2     9  14  R37221                IL-6 antagonist pe
  35    19   42.2     9  14  R50110                CDR (f).  Syntheti
  36    19   42.2     9  15  R73736                Antigen fragment 5
  37    19   42.2     9  16  R87430                Human MHC class II
  38    19   42.2     9  19  W70078                B. stearothermophi
  39    19   42.2     9  20  Y46676                Immunogenic peptid
  40    19   42.2     9  20  W96297                VatC peptide fragm
  41    19   42.2     9  21  B02515                WKY control sequen
  42    19   42.2     9  21  B02516                BN control sequenc
  43    19   42.2     9  21  Y80174                HLA-A2 restricted
  44    19   42.2     9  21  Y68806                Amino acid sequenc
  45    19   42.2     9  21  Y56575                Tyrosinase 2 deriv
  46    18   40.0     5  20  Y28182                Factor Xa inhibiti
  47    18   40.0     6  17  R90465                Antibody 561 displ
  48    18   40.0     6  20  Y55200                Anti CD34 antibody
  49    18   40.0     6  21  Y86946                Human haematopoiet
  50    18   40.0     7  19  Y21209                Human bcl2 proto-o
  51    18   40.0     7  20  Y41847                Rheumatoid arthrit
  52    18   40.0     7  20  Y17023                Heat shock protein
  53    18   40.0     8  16  R84478                Hepatitis C virus
  54    18   40.0     8  18  W36092                E. coli DNA polyme
  55    18   40.0     8  19  W57004                Enzyme inhibitor p
  56    18   40.0     8  20  Y16852                Heat shock protein
  57    18   40.0     8  20  W94102                VIF-derived HIV pr
  58    18   40.0     8  20  W82429                PVX-2 viral replic
  59    18   40.0     9  11  R07966                Tryptic fragment T
  60    18   40.0     9  15  R73856                Antigen fragment 1
  61    18   40.0     9  18  W38393                Synthetic pMEL17 p
  62    18   40.0     9  18  W38383                Synthetic pMEL17 p
  63    18   40.0     9  18  W15173                Chlamydia trachoma
  64    18   40.0     9  18  W15174                Chlamydia trachoma
  65    18   40.0     9  20  Y55366                HLA binding plu-1
  66    18   40.0     9  20  Y55448                HLA binding plu-1
  67    18   40.0     9  20  Y55529                HLA binding plu-1
  68    18   40.0     9  20  Y55623                HLA binding plu-1
  69    18   40.0     9  20  Y41907                Rheumatoid arthrit
  70    18   40.0     9  20  Y42029                Rheumatoid arthrit
  71    18   40.0     9  20  Y42038                Rheumatoid arthrit
  72    18   40.0     9  20  Y46029                Immunogenic peptid
  73    18   40.0     9  20  Y46442                Immunogenic peptid
  74    18   40.0     9  20  Y46500                Immunogenic peptid
  75    18   40.0     9  20  Y46533                Immunogenic peptid

                              ALIGNMENTS


RESULT   1
Y09528
ID   Y09528 standard; peptide; 9 AA.
XX
AC   Y09528;
```

```
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:4.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;    98WO-US19794.
XX
PR   10-OCT-1997;    97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;


  Query Match             100.0%;  Score 45;  DB 20;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGANINL 9
            |||||||||
Db        1 ylsganinl 9


RESULT   2
W39723
ID   W39723 standard; peptide; 9 AA.
XX
AC   W39723;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Human carcina-embryonic antigen (CEA) peptide (pos. 571-579).
XX
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
XX
OS   Homo sapiens.
XX
```

```
PN   WO9741440-A1.
XX
PD   06-NOV-1997.
XX
PF   28-APR-1997;    97WO-NL00229.
XX
PR   23-DEC-1996;    96EP-0203670.
PR   26-APR-1996;    96EP-0201145.
XX
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
PI   Kast WM,  Melief CJM,  Offringa R,  Toes REM,  Van Der Burg SH;
XX
DR   WPI; 1997-549891/50.
XX
PT   Method of selecting T cell peptide epitope(s) - by measuring the
PT   stability of HLA class I-peptide complexes on intact B cells
XX
PS   Example 3; Page 85; 109pp; English.
XX
CC   Peptides W39430-W39734 are used in a novel method for the selection of
CC   immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC   method involves the identification of peptide sequences capable of
CC   binding to an HLA (human leukocyte antigen) class I molecule and
CC   measuring the binding of this epitope peptide to the HLA class I peptide.
CC   The stability of binding of the peptide and MHC (major histocompatablilty
CC   complex) class I molecule is measured on intact human B cells carrying
CC   the MHC molecule at their cell surfaces. The method can be used to select
CC   peptide epitopes for generating vaccines against a disease associated
CC   with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC   especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC   immune responses. Peptide W39723 is derived from the human
CC   carcino-embryonic antigen (CEA) and has the ability to bind to the human
CC   MHC Class I allele HLA-A2.1.
XX
SQ   Sequence   9 AA;


  Query Match             95.6%;  Score 43;  DB 18;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGANINL 9
            ||||||:||
Db        1 ylsganlnl 9


RESULT   3
W77134
ID   W77134 standard; peptide; 9 AA.
XX
AC   W77134;
XX
DT   16-NOV-1998  (first entry)
XX
DE   CEA synthetic peptide epitope 1.
XX
KW   Tyrosinase; tyrosinase cytotoxic lymphocyte response;
KW   cytotoxic T lymphocyte; cysteine-depleted; melanoma.
XX
OS   Synthetic.
XX
PN   WO9833810-A2.
XX
PD   06-AUG-1998.
XX
PF   29-JAN-1998;    98WO-US01592.
XX
PR   30-JAN-1997;    97US-0037781.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
```

```
PI   Engelhard VH,  Hunt DF,  Kittlesen D,  Slingluff CL;
XX
DR   WPI; 1998-437388/37.
XX
PT   Disease specific immunogen  - comprises disease specific cytotoxic T
PT   lymphocyte epitope used to elicit melanoma specific CTL response
XX
PS   Disclosure; Page 27; 93pp; English.
XX
CC   The peptide epitope W77119-W77138 were created for human tumour-specific
CC   cytotoxic T lymphocyte response.  These peptides are are cysteine-
CC   depleted mutants of a native disease-specific CTL epitope.  The cysteine-
CC   depleted CTL epitopes elicit a stronger or more specific CTL response
CC   than the native epitope.  The epitopes can be used in a disease-specific
CC   immunogen to protect a mammal against disease in particular melanomas.
CC   The peptides may also be used to screen a sample for the presence of
CC   an antigen with the same epitope, or with a different cross-reactive
CC   epitope.
XX
SQ   Sequence   9 AA;

  Query Match             95.6%;  Score 43;  DB 19;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGANINL 9
            ||||||:||
Db        1 ylsganlnl 9


RESULT   4
W70045
ID   W70045 standard; peptide; 9 AA.
XX
AC   W70045;
XX
DT   22-OCT-1998  (first entry)
XX
DE   CEA derived HLA-A2.1 binding peptide 2 (residues 605-613).
XX
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW   antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW   viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW   fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9833888-A1.
XX
PD   06-AUG-1998.
XX
PF   30-JAN-1998;   98WO-US01959.
XX
PR   31-JAN-1997;   97US-0036696.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Celis E,  Sette A,  Sidney J,  Southwood S,  Tsai V;
XX
DR   WPI; 1998-437445/37.
XX
PT   Production of antigen-specific cytotoxic T cells - by incubating
PT   immunogenic peptide(s) from antigen that binds class I major
PT   histocompatibility complex molecules with pre-treated antigen
PT   presenting cells
XX
PS   Example 6; Page 75; 104pp; English.
XX
CC   Sequences shown in W70044 to W70052 represent peptides derived from
CC   carcinoembryonic antigen (CEA). The peptides can bind to a human
CC   leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method
CC   of invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC   vitro. The method comprises contacting immunogenic peptides from an
CC   antigen that binds class I major histocompatibility complex (MHC)
CC   molecules with antigen presenting cells (APCs) pretreated with
CC   pretreatment growth factors, and incubating the APCs with purified CD8
CC   cells in the presence of at least 2 incubation growth factors, thereby
CC   producing antigen-specific CTLs. A method for specifically killing
CC   target cells in a human patient is also provided which comprises
CC   obtaining a fluid sample containing CTLs from a patient, contacting the
CC   cytotoxic T cells with APCs pretreated with pre-treatment growth factors,
CC   where the APCs comprise class I MHC molecules. The pretreated APCs are
CC   incubated with the cytotoxic growth factors, thereby producing activated
CC   CTLs which are contacted with a carrier to form a composition. The
CC   composition can then be administered to the patient. The activated CTLs
CC   can be used for treating cancers, immune disorders, viral infections,
CC   AIDS, hepatitis, bacterial infection, fungal infection, malaria or
CC   tuberculosis.
XX
SQ   Sequence   9 AA;

  Query Match             95.6%;  Score 43;  DB 19;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGANINL 9
            ||||||:||
Db        1 ylsganlnl 9


RESULT   5
Y47655
ID   Y47655 standard; Peptide; 9 AA.
XX
AC   Y47655;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2266.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 118; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
```

```
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;

  Query Match            95.6%;  Score 43;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
           ||||||:||
Db       1 ylsganlnl 9

RESULT   6
Y09525
ID   Y09525 standard; peptide; 9 AA.
XX
AC   Y09525;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist CAP-1.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 1; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
```

```
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match            95.6%;  Score 43;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANINL 9
           ||||||:||
Db       1 ylsganlnl 9

RESULT   7
Y09527
ID   Y09527 standard; peptide; 9 AA.
XX
AC   Y09527;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:3.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match            88.9%;  Score 40;  DB 20;  Length 9;
```

```
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGANINL 9
            |||||:|||
Db        1 ylsgadinl 9


RESULT   8
Y54173
ID   Y54173 standard; peptide; 9 AA.
XX
AC   Y54173;
XX
DT   06-APR-2000  (first entry)
XX
DE   HLA binding peptide 1233.11 derived from source CEA.605V9.
XX
KW   Allele-specific binding motif; major histocompatibility complex; MHC;
KW   HLA; HLA-A2.1; cytotoxic T cell response; antigen; microbial infection;
KW   hepatitis; AIDS; malaria; condyloma acuminatum; cancer; prostate cancer;
KW   renal carcinoma; cervical carcinoma; lymphoma; tumour.
XX
OS   Unidentified.
XX
PN   WO9965522-A1.
XX
PD   23-DEC-1999.
XX
PF   17-JUN-1999;   99WO-US13789.
XX
PR   17-JUN-1998;   98US-0098584.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Sidney J,  Southwood S;
XX
DR   WPI; 2000-106018/09.
XX
PT   Novel HLA binding immunogenic peptides used to induce T cell activation
PT   and to induce an immune response  -
XX
PS   Claim 1; Page 32; 42pp; English.
XX
CC   Peptides Y54171-Y54236 represent immunogenic peptides comprising an
CC   allele-specific binding motif for the major histocompatibility complex
CC   (MHC) molecule HLA, e.g. HLA-A2.1. The peptides have conserved residues
CC   at certain positions such as positions 2 and 9. Also, the peptides do not
CC   comprise negative binding residues at other positions, such as positions
CC   1, 3, 6 and/or 7 (peptides 9 amino acids long) and at poisons 1, 3, 4,
CC   5, 7, 8 and/or 9 (peptides 10 amino acids long). The peptides are used to
CC   induce a cytotoxic T cell response to a preselected antigen. The method
CC   comprises contacting cytotoxic T cells from a patient (optionally
CC   expressing a specific MHC class I allele) with the present peptides.
CC   The peptides are used to treat and prevent microbial infection (e.g. in
CC   viral hepatitis B and C, human papilomavirus (HPV) infection, AIDS,
CC   cytomegalovirus (CMV), malaria, and condyloma acuminatum) and cancer
CC   (e.g.. prostate cancer, renal carcinoma, cervical carcinoma, lymphoma).
CC   Patients in the acute phase of infection can be treated with the
CC   peptides in conjunction with other treatments. The antigenic peptides
CC   may be used to elicit cytotoxic T lymphocytes (CTLs) ex vivo and in
CC   vivo. The resulting CTLs can be used to treat chronic infections (viral
CC   or bacterial) or tumours in patients that do not respond to conventional
CC   forms of therapy. The peptides may also be used to produce monoclonal
CC   antibodies, which are useful as potential diagnostic or therapeutic
CC   agents. The peptides may also be used as diagnostic reagents.
XX
SQ   Sequence   9 AA;

  Query Match             88.9%;  Score 40;  DB 21;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        1 YLSGANINL 9
            ||||||:|:
Db        1 ylsganlnv 9


RESULT   9
Y09526
ID   Y09526 standard; peptide; 9 AA.
XX
AC   Y09526;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:2.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             84.4%;  Score 38;  DB 20;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGANINL 9
            |||||::||
Db        1 ylsgadlnl 9


RESULT  10
W00680
ID   W00680 standard; peptide; 9 AA.
XX
AC   W00680;
```

```
XX
DT   01-MAY-1997  (first entry)
XX
DE   Peptide comprising residues 571-579 of Carcinoembryonic antigen.
XX
KW   Carcinoembryonic; antigen; human; cytotoxic T cell; pox virus;
KW   vector; epitope; determination; screening; tumour; treatment.
XX
OS   Homo sapiens.
XX
PN   WO9626271-A1.
XX
PD   29-AUG-1996.
XX
PF   13-FEB-1996;   96WO-US02156.
XX
PR   22-FEB-1995;   95US-0396385.
XX
PA   (THER-) THERION BIOLOGICS CORP.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Panicali D,  Schlom J,  Tsang KY;
XX
DR   WPI; 1996-402364/40.
XX
PT   Generation of human cytotoxic T-cells specific for CEA - useful in
PT   therapy, epitope mapping and drug screening
XX
PS   Claim 4; Page 57; 76pp; English.
XX
CC   Producing carcinoembryonic antigen (CEA) specific human cytotoxic T
CC   cells (CTC), comprises introducing a 1st pox virus vector, having
CC   at least 1 insertion site containing a DNA segment encoding a CEA
CC   peptide (i.e. the present peptide) to a host to stimulate CTC
CC   production, and at least 1 periodic interval after that, contacting
CC   the host with an additional antigen. The CEA specific CTC can be
CC   used to determine the CTC eliciting epitope of CEA, and to screen
CC   for compounds which enhance the ability of the antigen to create a
CC   CTC response. A host with a CEA expressing tumour can be treated by
CC   introducing the CTC to the host, and at least 1 periodic interval
CC   after that introducing a CEA peptide, i.e. the present peptide.
CC   The present peptide is positive for binding to HLA-A2, and scored
CC   561 and 806 in T2 cell binding assays, where the binding of an
CC   appropriate peptide results in the upregulation of surface HLA-A2
CC   on the T2 cells, which can be quantified via FACScan using an
CC   anti-HLA-A2 antibody (background 280 and 300).
XX
SQ   Sequence   9 AA;

  Query Match             80.0%;  Score 36;  DB 17;  Length 9;
  Best Local Similarity   87.5%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 LSGANINL 9
          |||||:||
Db      2 lsganlnl 9

RESULT  11
Y09529
ID   Y09529 standard; peptide; 9 AA.
XX
AC   Y09529;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:5.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
```

```
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             75.6%;  Score 34;  DB 20;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANINL 9
          ||||| :||
Db      1 ylsgaclnl 9

RESULT  12
W00690
ID   W00690 standard; peptide; 9 AA.
XX
AC   W00690;
XX
DT   01-MAY-1997  (first entry)
XX
DE   NCA analogue of residues 571-579 of carcinoembryonic antigen.
XX
KW   Carcinoembryonic; antigen; epitope; NCA; analogue.
XX
OS   Homo sapiens.
XX
PN   WO9626271-A1.
XX
PD   29-AUG-1996.
XX
PF   13-FEB-1996;   96WO-US02156.
XX
PR   22-FEB-1995;   95US-0396385.
XX
PA   (THER-) THERION BIOLOGICS CORP.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
```

```
PI   Panicali D,  Schlom J,  Tsang KY;
XX
DR   WPI; 1996-402364/40.
XX
PT   Generation of human cytotoxic T-cells specific for CEA - useful in
PT   therapy, epitope mapping and drug screening
XX
PS   Example 2; Page 60; 76pp; English.
XX
CC   The present peptide is negative for binding to HLA-A2, and scored
CC   252 and 225 in T2 cell binding assays, where the binding of an
CC   appropriate peptide results in the upregulation of surface HLA-A2
CC   on the T2 cells, which can be quantified via FACScan using an
CC   anti-HLA-A2 antibody (background 280 and 300).
XX
SQ   Sequence   9 AA;


  Query Match            60.0%;  Score 27;  DB 17;  Length 9;
  Best Local Similarity   55.6%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 YLSGANINL 9
          |  | |:||
Db      1 yrpgenlnl 9


RESULT  13
W54298
ID   W54298 standard; Peptide; 9 AA.
XX
AC   W54298;
XX
DT   28-AUG-1998  (first entry)
XX
DE   Human cytohesin-1 PH domain randomised AB loop (clone 7).
XX
KW   Pleckstrin homology domain; PH domain; peptide library;
KW   functional molecular surface; protein structural template;
KW   vaccine; gene therapy; cytohesin 1; human.
XX
OS   Synthetic.
XX
PN   WO9745538-A1.
XX
PD   04-DEC-1997.
XX
PF   30-MAY-1997;   97WO-EP02840.
XX
PR   31-MAY-1996;   96EP-0108776.
XX
PA   (MEDI-) MEDIGENE AG.
XX
PI   Bruhn H,  Funk M,  Henkel T,  Steipe B;
XX
DR   WPI; 1998-230215/20.
DR   N-PSDB; V26501.
XX
PT   Vectors used to produce PH domain-like peptide libraries - which are
PT   screened for therapeutically useful peptide(s), e.g. to produce
PT   vaccines
XX
PS   Example 5; Fig 12; 137pp; English.
XX
CC   This is a randomised AB loop encoded by clone 7 (see V26501) of a
CC   large peptide library (2 million clones) of synthetic Pleckstrin
CC   homology (PH) domains with randomised discontinuous surface
CC   epitopes.  The progenitor AB sequence comprises amino acid residues
CC   16-21 of a stabilised synthetic PH domain (see W54310) of human
CC   cytohesin 1.  The peptide library was generated by randomisation of
CC   AB and CD loop regions using randomised oligonucleotide primers
CC   (see W26493-94).  The randomised AB and CD loop regions of 9 clones
CC   (see W54292-309) of the peptide library are provided.  Randomisation
```

```
CC   did not compromise the structural integrity and folding stability of
CC   the progenitor domain.  The invention provides vectors that are used
CC   for the production of PH domain-like peptide libraries, which can
CC   be screened to identify peptides that have desirable properties,
CC   especially novel binding or catalytic properties, and which may
CC   be of use in research or therapy, or as vaccines.  Novel synthetic
CC   protein structural templates for the generation, screening and
CC   evolution of functional molecular surfaces are provided.
XX
SQ   Sequence   9 AA;


  Query Match            51.1%;  Score 23;  DB 19;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 LSGANIN 8
          |||  :|
Db      1 lsggrvn 7


RESULT  14
W19861
ID   W19861 standard; peptide; 9 AA.
XX
AC   W19861;
XX
DT   24-FEB-1998  (first entry)
XX
DE   Fragment of enzyme capable of catalysing conversion of DNA to CO-DNA.
XX
KW   CO-DNA; cell division; cell proliferation; tumour; lysilendopeptidase;
KW   N-terminal fragment; V8 proteinase digestion.
XX
OS   Gallus domesticus.
XX
PN   WO9720919-A1.
XX
PD   12-JUN-1997.
XX
PF   27-NOV-1996;   96WO-IB01323.
XX
PR   06-DEC-1995;   95US-0567934.
XX
PA   (KAGA/) KAGAWA H.
PA   (KAGA/) KAGAWA K.
PA   (TOKI/) TOKIMATSU H.
XX
PI   Kagawa H,  Kagawa K,  Tokimatsu H;
XX
DR   WPI; 1997-319767/29.
XX
PT   New enzyme which converts DNA to CO-DNA - used to reduce or
PT   eliminate the ability of a cell to divide, for inhibiting cell
PT   proliferation and treating tumours
XX
PS   Claim 3; Page 17; 35pp; English.
XX
CC   This sequence represents a V8 proteinase digestion product of an
CC   isolated pure enzyme which is capable of catalysing the conversion of
CC   DNA to CO-DNA. CO-DNA is a form of DNA in which a carbonyl group is
CC   attached to the 1' carbon of the sugar constituent of the DNA. The enzyme
CC   can be used to convert DNA to CO-DNA which reduces or eliminates the
CC   ability of a cell to divide. The enzyme can be used for inhibiting cell
CC   proliferation and for treating tumours.
XX
SQ   Sequence   9 AA;


  Query Match            48.9%;  Score 22;  DB 18;  Length 9;
  Best Local Similarity   42.9%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy       2 LSGANIN 8
           : |||::
Db       3 idganvh 9


RESULT  15
R09414
ID   R09414 standard; Peptide; 6 AA.
XX
AC   R09414;
XX
DT   23-AUG-1990  (first entry)
XX
DE   LFA-1 alpha subunit polypeptide (k).
XX
KW   Lymphocyte function associated antigen; inflammation; metastasis.
XX
PN   EP362526-A.
XX
PD   11-APR-1990.
XX
PF   17-AUG-1989;   89EP-0115160.
XX
PR   23-AUG-1988;   88US-0235227.
PR   09-MAR-1989;   89US-0321017.
XX
PA   (DANA-) DANA FARBER CANCER.
XX
PI   Springer TA,  Larson R;
XX
DR   WPI; 1990-108985/15.
XX
PT   Pure alpha subunit of lymphocyte function associated antigen -
PT   and encoding DNA sequences, useful eg for suppressing
PT   inflammation or metastasis
XX
PS   Claim 6; Page 19; 27pp; English.
XX
CC   The alpha-subunit (a-SU), contg. al least one of the polypeptides
CC   given in R09404-417, can bind to ICAM-1 (or other natural
CC   ligands) on the surface of cells, and can associate with the beta-SU
CC   to form a heterodimer (also able to bind to ICAM-I).  a-SU, and its
CC   derivs., are useful in suppressing inflammation, metastasis and
CC   growth of a-SU expressing tumour cells and is used in the treatment
CC   of viral infections.
CC   The pref. dose is 1 pg - 10 mg/kg.
XX
SQ   Sequence   6 AA;


  Query Match             46.7%;  Score 21;  DB 11;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
             ||||
Db       2 ylsg 5


RESULT  16
R88476
ID   R88476 standard; peptide; 6 AA.
XX
AC   R88476;
XX
DT   30-AUG-1996  (first entry)
XX
DE   Internal tryptic peptide from Tre6P synthase (peak 29) #2.
XX
KW   Tryptic peptide; trehalose-6-phosphate synthase; M. smegmatis; probe;
KW   trehalose; transgenic plant; heparin-activated; preservation; food;
KW   antigenic determinant; yeast; TSP1; fruit; berry; puree; jelly; jam.
XX
```

```
OS   Mycobacterium smegmatis.
XX
PN   WO9600789-A1.
XX
PD   11-JAN-1996.
XX
PF   29-JUN-1995;   95WO-FI00377.
XX
PR   29-JUN-1994;   94FI-0003133.
XX
PA   (ALKO-) ALKO GROUP LTD.
XX
PI   Holmstrom K,  Londesborough J,  Mandal A,  Mantyla E;
PI   Palva ET,  Tunnela O, Welin B;
XX
DR   WPI; 1996-077499/08.
XX
PT   New transgenic plants with increase trehalose contents - prepd. by
PT   transforming plants with a trehalose-6-phosphate synthase gene fused
PT   to a non-constitutive promoter
XX
PS   Example 6; Page 36; 55pp; English.
XX
CC   The sequences given in R88473-80 are internal tryptic peptides
CC   derived from trehalose-6-phosphate (Tre6P) synthase from M.
CC   smegmatis.  Tre6P is the key enzyme in the synthesis of trehalose
CC   via Tre6P.  The aim of the invention is to produce a transgenic
CC   plant with increase trehalose content.  Tre6P in M. smegmatis is
CC   heparin-activated and was isolated  and purified.   These peptides
CC   were derived from a protein which was purified with a mol. wt. of 55
CC   kD which  shared antigenic determinants with the yeast Tre6P
CC   synthase protein.  Using these peptides probes may be designed for
CC   the isolation of the Tre6P gene (TSP1) for the production of the
CC   transgenic plants.  The trehalose may be isolated from the transgenic
CC   plants and used in bulk preparation including the preservation of the
CC   flavour and structure of food stuffs during drying.  Fruits and berries
CC   form the transformed plants may be processed into purees, jellies and
CC   jams which have a fresher and richer flavour due to the increased
CC   trehalose content.
XX
SQ   Sequence   6 AA;


  Query Match             46.7%;  Score 21;  DB 17;  Length 6;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGA 5
             || ||
Db       1 ylega 5


RESULT  17
R29031
ID   R29031 standard; Protein; 8 AA.
XX
AC   R29031;
XX
DT   20-APR-1993  (first entry)
XX
DE   Peptide encoded by ant-active toxin gene probe.
XX
KW   Toxin protein; ant.
XX
OS   Synthetic.
XX
FH   Key                Location/Qualifiers
FT   Misc-difference 1
FT                      /label= Arg or Lys
FT   Misc-difference 3
FT                      /label= Trp or Tyr
FT   Misc-difference 4
FT                      /label= Ile or Leu
```

```
FT   Misc-difference 7
FT                   /label= Ala or Asn
FT   Misc-difference 8
FT                   /label= Asn or Gln
XX
PN   WO9220802-A.
XX
PD   26-NOV-1992.
XX
PF   22-MAY-1992;   92WO-US04316.
XX
PR   22-MAY-1991;   91US-0703977.
PR   25-NOV-1991;   91US-0797645.
PR   12-MAY-1992;   92EP-0304228.
XX
PA   (MYCO ) MYCOGEN CORP.
XX
PI   Kennedy MK,  Meier H,  Payne JM,  Randall JB,  Uick HJ;
XX
DR   WPI; 1992-415780/50.
DR   N-PSDB; Q31414, Q31415.
XX
PT   Toxin proteins isolated from Bacillus thuringiensis - for controlling
PT   amts. e.g. fire, carpenter, argentine and pharaoh ants
XX
PS   Disclosure; Page 56; 71pp; English.
XX
CC   The peptide (or point mutation variants as shown in the features)
CC   is encoded by nucleotide probes used in the rapid identification of
CC   Bacillus thuringiensis ant-active toxin genes.
XX
SQ   Sequence   8 AA;


  Query Match             46.7%;  Score 21;  DB 13;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGAN 6
           :::|||
Db       3 wingan 8


RESULT  18
R78208
ID   R78208 standard; Peptide; 8 AA.
XX
AC   R78208;
XX
DT   22-FEB-1996  (first entry)
XX
DE   B. thuringiensis endotoxin derived peptide.
XX
KW   Endotoxin; probes; microbes; peptide; detection; nematode.
XX
OS   Bacillus thuringiensis.
XX
PN   US5430137-A.
XX
PD   04-JUL-1995.
XX
PF   25-OCT-1989;   89US-0427068.
XX
PR   30-OCT-1992;   92US-0968781.
PR   25-OCT-1989;   89US-0427068.
PR   26-JUL-1991;   91US-0737569.
XX
PA   (MYCO ) MYCOGEN CORP.
XX
PI   Gaertner FH,  Narva KE,  Schnepf HE,  Schwab GE,  Sick AJ;
PI   Thompson M;
XX
DR   WPI; 1995-245777/32.
```

```
XX
PT   Nucleotide sequence used as probes to identify Bacillus
PT   thuringiensis - are derived from the B.thuringiensis endotoxin
PT   genes, for identifying microbes which encode toxins
XX
PS   Claim 1; Columns 41-42; 30pp; English.
XX
CC   The nucleotide sequence which encodes R78208, a B. thuringiensis
CC   (B.t.) endotoxin derived peptide, specifies the degenerate
CC   probe Q94857. The probe can be used for the detection
CC   of endotoxin producing B.t. microbes. The probe aids in the search
CC   for useful microbes hosting toxin encoding genes, specifically from
CC   nematode species.
XX
SQ   Sequence   8 AA;


  Query Match             46.7%;  Score 21;  DB 16;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGAN 6
           :::|||
Db       3 wingan 8


RESULT  19
R73808
ID   R73808 standard; peptide; 9 AA.
XX
AC   R73808;
XX
DT   19-JUN-1995  (first entry)
XX
DE   Antigen fragment 124, from LCMV has binding affinity for HLA-2.1.
XX
KW   antigen; epitope; immunogenic target protein; PSA; HBVc; HBVs; EBV;
KW   HIV1; plasma specific antigen; hepatitis B virus; Epstein Barr;
KW   human immunodeficiency virus; human papilloma virus; p53; c-ERB2;
KW   MAGE-1; melanoma antigen-1; core antigen; surface antigen;
KW   pharmaceutical composition; in vivo; ex vivo; therapeutic;
KW   diagnostic; MHC class I molecule; major histocompatability complex;
KW   HLA-A2.1; 9mer; 10mer; anchor; human leukocyte antigen; PLP; 8mer;
KW   algorithm prediction; MBP; CMV; cytomegalovirus; HSV;
KW   herpes simplex virus; influenza A; M1; LCMV.
XX
OS   Synthetic.
XX
PN   WO9420127-A.
XX
PD   15-SEP-1994.
XX
PF   04-MAR-1994;   94WO-US02353.
XX
PR   05-MAR-1993;   93US-0027146.
PR   04-JUN-1993;   93US-0073205.
PR   29-NOV-1993;   93US-0159184.
XX
PA   (CYTE-) CYTEL CORP.
XX
PI   Grey HM,  Kast WM,  Sette A,  Sidney J;
XX
DR   WPI; 1994-302678/37.
XX
PT   Immunogenic peptide(s) having an HLA-A2.1 binding motif - used
PT   for treatment or prophylaxis of cancer, virus infection or
PT   autoimmune diseases.
XX
PS   Disclosure; Page 84; 138pp; English.
XX
CC   R73685-876 are potential peptide binders of HLA-A2.1 motif. Using
CC   motifs disclosed in the invention, these peptides were screened for
CC   further motifs. Only peptides with binding affinity of at least 1%
```

```
CC   (binding affinity is expressed as an IC50 value) as compared to the
CC   standard peptide (R71293) in assays. This peptide from LCMV (sic) Np
CC   antigen has a binding value of 0.0280. The peptides of the invention
CC   can induce cytotoxic T lymphocytes which can react with target cells.
CC   They can be used for the treatment or prophlyaxis of cancer, eg.
CC   prostate cancer or lymphoma, etc.
XX
SQ   Sequence   9 AA;

  Query Match             46.7%;  Score 21;  DB 15;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 LSGANINL 9
              :|| | :|
Db          2 isgynfsl 9


RESULT  20
W49339
ID   W49339 standard; peptide; 9 AA.
XX
AC   W49339;
XX
DT   05-JUN-1998  (first entry)
XX
DE   Human leucocyte antigen DQ4 binding peptide #230.
XX
KW   Human leucocyte antigen; HLA-DQ4; combinatorial library;
KW   autoimmune disease; chronic articular rheumatism.
XX
OS   Synthetic.
XX
PN   JP08151396-A.
XX
PD   11-JUN-1996.
XX
PF   28-NOV-1994;   94JP-0292657.
XX
PR   28-NOV-1994;   94JP-0292657.
XX
PA   (TEIJ ) TEIJIN LTD.
XX
DR   WPI; 1996-329479/33.
XX
PT   HLA-binding oligopeptide and an immuno:regulator contg it - used in
PT   the treatment of auto:immune disease
XX
PS   Claim 4; Page 31; 61pp; Japanese.
XX
CC   This peptide is an example of a peptide which binds to a human leucocyte
CC   antigen HLA-DQ4 molecule.  The peptide was isolated from a phagemid
CC   combinatorial library comprising the sequence V05953, by screening with
CC   an HLA-DQ4 molecule.  The peptide is used for the treatment of autoimmune
CC   disease, or especially for treatment of viral diseases.
XX
SQ   Sequence   9 AA;

  Query Match             46.7%;  Score 21;  DB 17;  Length 9;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 GANIN 8
              ||||:
Db          1 ganis 5


RESULT  21
W49340
ID   W49340 standard; peptide; 9 AA.
XX
AC   W49340;
XX
DT   05-JUN-1998  (first entry)
XX
DE   Human leucocyte antigen DQ4 binding peptide #231.
XX
KW   Human leucocyte antigen; HLA-DQ4; combinatorial library;
KW   autoimmune disease; chronic articular rheumatism.
XX
OS   Synthetic.
XX
PN   JP08151396-A.
XX
PD   11-JUN-1996.
XX
PF   28-NOV-1994;   94JP-0292657.
XX
PR   28-NOV-1994;   94JP-0292657.
XX
PA   (TEIJ ) TEIJIN LTD.
XX
DR   WPI; 1996-329479/33.
XX
PT   HLA-binding oligopeptide and an immuno:regulator contg it - used in
PT   the treatment of auto:immune disease
XX
PS   Claim 4; Page 31; 61pp; Japanese.
XX
CC   This peptide is an example of a peptide which binds to a human leucocyte
CC   antigen HLA-DQ4 molecule.  The peptide was isolated from a phagemid
CC   combinatorial library comprising the sequence V05953, by screening with
CC   an HLA-DQ4 molecule.  The peptide is used for the treatment of autoimmune
CC   disease, or especially for treatment of viral diseases.
XX
SQ   Sequence   9 AA;

  Query Match             46.7%;  Score 21;  DB 17;  Length 9;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 GANIN 8
              ||||:
Db          1 ganis 5


RESULT  22
Y83831
ID   Y83831 standard; peptide; 6 AA.
XX
AC   Y83831;
XX
DT   05-JUL-2000  (first entry)
XX
DE   Pyrrolobenzodiazepine-peptide #6 from combinatorial library screen.
XX
KW   Antibacterial; antiparasitic; cytostatic; pyrrolobenzodiazepine;
KW   DNA minor groove; combinatorial peptide library screening; virucide;
KW   cytotoxic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "conjugated to pyrrolobenzodiazepine group"
XX
PN   WO200012506-A2.
XX
PD   09-MAR-2000.
XX
PF   27-AUG-1999;   99WO-GB02836.
XX
PR   27-AUG-1998;   98GB-0018730.
```

```
XX
PA   (UYPO-) UNIV PORTSMOUTH HIGHER EDUCATION CORP.
XX
PI   Thurston DE,  Howard PW;
XX
DR   WPI; 2000-237841/20.
XX
PT   New pyrrolobenzodiazepine derivatives used in the preparation of
PT   combinatorial libraries of further pyrrolobenzodiazepine derivatives
PT   -
XX
PS   Example 8; Page 72; 158pp; English.
XX
CC   The invention relates to novel pyrrolobenzodiazepine (PBD) derivatives
CC   which are compounds that bind to the minor groove of double stranded
CC   nucleic acid sequences, especially at the sequence Purine-Guanine-Purine
CC   where Purine is selected from adenine and guanine. The PBD's can be used
CC   in combinatorial peptide library screening for binding interactions with
CC   nucleic acids. The PBD's can be used in cytotoxic, antibiotic,
CC   antiparasitic and antiviral compositions and also in methods of diagnosis
CC   and in methods of target validation in functional genomics. The peptides
CC   Y83826-Y83833 represent PBD-peptides isolated froma combinatorial peptide
CC   library screen on the target nucleic acid sequence Z99178.
XX
SQ   Sequence   6 AA;


  Query Match             44.4%;  Score 20;  DB 21;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 GANI 7
              ||||
Db          2 gani 5


RESULT  23
Y41846
ID   Y41846 standard; Peptide; 7 AA.
XX
AC   Y41846;
XX
DT   09-DEC-1999  (first entry)
XX
DE   Rheumatoid arthritis diagnostic protein isoform 17 peptide #1.
XX
KW   Human; rheumatoid arthritis; RA; diagnosis; RPI; RADF; detection;
KW   rheumatoid arthritis diagnostic feature; ERPI; synovial fluid;
KW   rheumatoid arthritis diagnostic protein isoform; screening;
KW   expression reference protein isoform; prognosis.
XX
OS   Homo sapiens.
XX
PN   WO9947925-A2.
XX
PD   23-SEP-1999.
XX
PF   15-MAR-1999;   99WO-GB00763.
XX
PR   13-MAR-1998;   98GB-0005477.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Parekh RB,  Patel TP,  Townsend RR;
XX
DR   WPI; 1999-571871/48.
XX
PT   Diagnosis of human rheumatoid arthritis by two-dimensional
PT   electrophoresis -
XX
PS   Claim 20; Page 150; 157pp; English.
XX
CC   A method has been developed for the diagnosis of human rheumatoid
```

```
CC   arthritis (RA) using two-dimensional electrophoresis to generate a
CC   two-dimensional array of features. The method can be used for screening,
CC   diagnosis and prognosis of RA in a subject or for monitoring the effect
CC   of an anti-RA drug or therapy administered to a subject. The method
CC   comprises: (a) analysing a sample of serum or plasma and optionally
CC   synovial fluid by two-dimensional electrophoresis, to generate a two-
CC   dimensional array of features; (b) identifying at least one chosen
CC   feature whose relative abundance correlates with the presence or absence
CC   of RA; and (c) comparing the abundance of each chosen feature in the
CC   sample with the abundance of that chosen feature in serum or plasma from
CC   one or more persons without RA, where the relative abundance of the
CC   chosen feature or features in the sample indicates the presence or
CC   absence of RA in the subject. The method can also be used in clinical
CC   studies for testing drugs for therapy of RA, for purification of RA-
CC   diagnostic protein isoforms (RPIs), and for production of antibodies to
CC   RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify
CC   compounds that promote or inhibit their activity, which are then used as
CC   RA drugs. Nucleic acid encoding RADFs can be used in gene therapy
CC   protocols. Y41844 to Y42100 represent RPI peptides, Y42101 to Y42103
CC   represent expression reference protein isoform peptides and Z25066 to
CC   Z25068 represent degenerate probes for RPIs, which are all used in
CC   the exemplification of the present invention.
XX
SQ   Sequence   7 AA;

  Query Match             44.4%;  Score 20;  DB 20;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 SGANIN 8
              |||:|:
Db          2 sgadis 7


RESULT  24
Y54382
ID   Y54382 standard; Peptide; 7 AA.
XX
AC   Y54382;
XX
DT   06-APR-2000  (first entry)
XX
DE   Amino acid sequence of a fragment of a heliomycin peptide.
XX
KW   Heliomycin; polycysteine peptide; antifungal; antibacterial;
KW   fungal infection; plant protection; animal protection;
KW   transgenic plant; fungi; bacteria; Cercospora beticola;
KW   Cladosporium herbarum; Fusarium culmorum; F. oxysporum;
KW   Phytophthora cinnamoni.
XX
OS   Heliothis virescens.
XX
PN   FR2777568-A1.
XX
PD   22-OCT-1999.
XX
PF   15-APR-1998;   98FR-0004933.
XX
PR   15-APR-1998;   98FR-0004933.
XX
PA   (RHON ) RHONE-POULENC AGROCHIMIE.
XX
DR   WPI; 2000-108532/10.
XX
PT   New polycysteine peptides, designated heliomycin, with antifungal or
PT   antibacterial activity, useful in medicine or plant protection  -
XX
PS   Claim 6; Page 37; 46pp; French.
XX
CC   Y54379-82 represent fragments of the generic heliomycin peptide
CC   given in Y54378. Heliomycin is a polycysteine peptide with antifungal
CC   and antibacterial activity, isolated from the lepidopteron Heliothis
```

```
CC   virescens. The heliomycin peptides are used as pharmaceuticals for
CC   treating or preventing fungal infections in humans and animals. They
CC   may also be used as antifungal agents for plant protection. Nucleic
CC   acid encoding the heliomycin peptides is used to generate transgenic
CC   plants that are resistant to some fungi and bacteria (specifically
CC   Cercospora beticola, Cladosporium herbarum, Fusarium culmorum,
CC   F. oxysporum and Phytophthora cinnamoni), and for recombinant
CC   production of the peptides.
XX
SQ   Sequence   7 AA;

  Query Match             44.4%;  Score 20;  DB 21;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        3 SGANIN 8
            | ||:|
Db        2 sfanvn 7


RESULT  25
R69993
ID   R69993 standard; peptide; 9 AA.
XX
AC   R69993;
XX
DT   19-OCT-1995  (first entry)
XX
DE   Nonameric mimotope 24 used to obtain highly specific antibodies.
XX
KW   mimotope; antibody; production; high specificity; detection;
KW   immunoassay; high performance liquid chromatography.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   9
FT                   /note= "Arg-NH2"
XX
PN   US5384263-A.
XX
PD   24-JAN-1995.
XX
PF   13-OCT-1987;   87US-0108130.
XX
PR   11-OCT-1988;   88US-0255906.
PR   13-OCT-1987;   87US-0108130.
PR   04-JUN-1993;   93US-0072190.
XX
PA   (TERR-) TERRAPIN TECHNOLOGIES INC.
XX
PI   Kauvar LM;
XX
DR   WPI; 1995-105497/14.
XX
PT   Producing antibodies with high specificity and affinity for an
PT   analyte  - by immunisation with selected mimotope, also analyte
PT   detection kits, useful for immunoassay of materials usually
PT   analysed by HPLC
XX
PS   Example 5; Fig 6; 25pp; English.
XX
CC   R69970-993 are nonameric mimotopes designed to show high diversity
CC   in hydrophobic moment and hydrophobic index, as well as charge
CC   distribution and size. The mimotopes are used in the method of the
CC   invention to obtain antibodies specifically and strongly reactive
CC   with a desired analyte. 16 of the peptides were tested for ability
CC   to bind the murine antibody Mab33-6, arbitrarily chosen, and known
CC   to bind to the peptides MB3 and MB4. 3 of the 16 peptides successfully
CC   bound Mab33-6. The mimotope is obtd. by reacting a panel of starting
CC   antibodies (Abs) representative of the resting B cell repertoire of a
CC   mammal with an analyte (so as to identify analyte-reacting Abs) and
```

```
CC   then reacting each of a panel of candidate mimotopes representative
CC   of a random set of 3D contours with the analyte-reacting Abs. A
CC   subject is immunised with one or more mimotopes identified and the
CC   product Abs are recovered from the serum of the subject.
XX
SQ   Sequence   9 AA;

  Query Match             44.4%;  Score 20;  DB 16;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 SGAN 6
            ||||
Db        5 sgan 8

RESULT  26
W54299
ID   W54299 standard; Peptide; 9 AA.
XX
AC   W54299;
XX
DT   28-AUG-1998  (first entry)
XX
DE   Human cytohesin-1 PH domain randomised AB loop (clone 8).
XX
KW   Pleckstrin homology domain; PH domain; peptide library;
KW   functional molecular surface; protein structural template;
KW   vaccine; gene therapy; cytohesin 1; human.
XX
OS   Synthetic.
XX
PN   WO9745538-A1.
XX
PD   04-DEC-1997.
XX
PF   30-MAY-1997;   97WO-EP02840.
XX
PR   31-MAY-1996;   96EP-0108776.
XX
PA   (MEDI-) MEDIGENE AG.
XX
PI   Bruhn H,  Funk M,  Henkel T,  Steipe B;
XX
DR   WPI; 1998-230215/20.
DR   N-PSDB; V26502.
XX
PT   Vectors used to produce PH domain-like peptide libraries - which are
PT   screened for therapeutically useful peptide(s), e.g. to produce
PT   vaccines
XX
PS   Example 5; Fig 12; 137pp; English.
XX
CC   This is a randomised AB loop encoded by clone 8 (see V26502) of a
CC   large peptide library (2 million clones) of synthetic Pleckstrin
CC   homology (PH) domains with randomised discontinuous surface
CC   epitopes.  The progenitor AB sequence comprises amino acid residues
CC   16-21 of a stabilised synthetic PH domain (see W54310) of human
CC   cytohesin 1.  The peptide library was generated by randomisation of
CC   AB and CD loop regions using randomised oligonucleotide primers
CC   (see W26493-94).  The randomised AB and CD loop regions of 9 clones
CC   (see W54292-309) of the peptide library are provided.  Randomisation
CC   did not compromise the structural integrity and folding stability of
CC   the progenitor domain.  The invention provides vectors that are used
CC   for the production of PH domain-like peptide libraries, which can
CC   be screened to identify peptides that have desirable properties,
CC   especially novel binding or catalytic properties, and which may
CC   be of use in research or therapy, or as vaccines.  Novel synthetic
CC   protein structural templates for the generation, screening and
CC   evolution of functional molecular surfaces are provided.
XX
SQ   Sequence   9 AA;
```

```
  Query Match             44.4%;  Score 20;  DB 19;  Length 9;
  Best Local Similarity   42.9%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LSGANIN 8
           |:|  :|
Db       1 laggkvn 7


RESULT  27
W40267
ID   W40267 standard; Protein; 9 AA.
XX
AC   W40267;
XX
DT   16-JUN-1998  (first entry)
XX
DE   K. oxytoca R-specific amidohydrolase peptide T5.
XX
KW   R-specific amidohydrolase; hydrolysis; racemate; nitrogen source;
KW   3,3,3-trifluoro-2-hydroxy-2-methyl propionamide; isoform.
XX
OS   Klebsiella oxytoca.
XX
PN   WO9801568-A2.
XX
PD   15-JAN-1998.
XX
PF   10-JUL-1997;   97WO-EP03670.
XX
PR   03-MAR-1997;   97CH-0000500.
PR   10-JUL-1996;   96CH-0001723.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Brieden W,  Naughton A,  Robins K,  Shaw N,  Tinschert A;
PI   Zimmermann T;
XX
DR   WPI; 1998-101063/09.
XX
PT   Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation
PT   - by stereoselective hydrolysis of corresponding racemic amide using
PT   microorganism or derived enzyme, used as drug intermediate
XX
PS   Example 10.2; Page 29; 68pp; German.
XX
CC   W40264-W40275 are peptide fragments of an R-specific amidohydrolase
CC   isolated from Klebsiella oxytoca strain PRS1. This enzyme allows the
CC   microorganism to utilise 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
CC   as its sole nitrogen source. This amidohydrolase is used in a process for
CC   preparing (R)-isoforms of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
CC   which is cheaper than prior art optical resolution of the racemate using
CC   dimethoxy strychnine or (S)-(-)- alpha -methylbenzylamine.
XX
SQ   Sequence   9 AA;


  Query Match             44.4%;  Score 20;  DB 19;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 YLSGANIN 8
           |  || :|
Db       1 ytvgamln 8


RESULT  28
Y23411
ID   Y23411 standard; Peptide; 6 AA.
XX
AC   Y23411;
XX
DT   02-SEP-1999  (first entry)
XX
DE   V beta 6 clone found in MS patients after vaccination with TCR.
XX
KW   Vaccine; T cell receptor; TCR; T cell; V beta 6.2/3; V beta 6/5;
KW   V beta 6.7; V beta 2; V beta 5/1; V beta 7; V beta 13; V beta 6;
KW   multiple sclerosis.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9927957-A1.
XX
PD   10-JUN-1999.
XX
PF   03-DEC-1997;   97WO-US23147.
XX
PR   03-DEC-1997;   97WO-US23147.
XX
PA   (IMMU-) IMMUNE RESPONSE CORP.
PA   (KIMM-) KIMMEL CANCER CENT SIDNEY.
XX
PI   Brostoff SW,  Carlo DJ,  Gold DP,  Smith LR,  Wilson DB;
XX
DR   WPI; 1999-404801/34.
XX
PT   T0 cell receptor peptide-derived vaccines
XX
PS   Example 11; Page 85; 104pp; English.
XX
CC   The specification describes vaccines which comprise immunologically
CC   effective amounts of T cell receptor (TCR) peptides. The TCRs are
CC   present on the surface of T cells. The TCRs are chosen from V beta 6.2/3,
CC   V beta 6/5, V beta 6.7, V beta 2, V beta 5/1, V beta 7 or V beta 13.
CC   The V beta TCR peptide-based vaccines are useful for prevention or
CC   treatment of multiple sclerosis (MS). The presence of V beta 6.7 appears
CC   to be particularly associated with multiple sclerosis and can be used
CC   to determine an individual's susceptibility to multiple sclerosis.
CC   Vaccinating, rather than passively administering heterologous antibodies,
CC   allows the host's own immune system to mobilize and suppress auto
CC   aggressive T cells.  Therefore, the suppression is persistent and may
CC   involve any and all immunological mechanisms in effecting that
CC   suppression.  Such a multi-faceted response is more effective than
CC   the uni-dimensional suppression achieved by passive administration of
CC   monoclonal antibodies or extant-derived regulatory T cell clones.
CC   Y23387-Y23480 represent peptides derived from TCR V beta 6 clones
CC   found in the cerebrospinal fluid (CFS) of MS patients, after vaccination
CC   with V beta 6.
XX
SQ   Sequence   6 AA;


  Query Match             42.2%;  Score 19;  DB 20;  Length 6;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LSGANI 7
           |:| |:
Db       1 lagtnl 6


RESULT  29
R75676
ID   R75676 standard; peptide; 7 AA.
XX
AC   R75676;
XX
DT   12-JAN-1996  (first entry)
XX
DE   Streptokinase-binding plasmin B chain peptide.
XX
KW   Streptokinase; plasmin B; affinity chromatography; purification;
```

```
KW   isolation.
XX
OS   Homo sapiens.
XX
PN   GB2284422-A.
XX
PD   07-JUN-1995.
XX
PF   29-NOV-1994;   94GB-0024112.
XX
PR   29-NOV-1993;   93GB-0024473.
XX
PA   (OXFO-) OXFORD RES SUPPORT CO LTD.
XX
PI   Cederholm-Williams SA;
XX
DR   WPI; 1995-196178/26.
XX
PT   New peptide sequences from plasmin B chain - able to bind other
PT   proteins esp. streptokinase, useful for affinity purifcn.
XX
PS   Claim 3; Page 6; 14pp; English.
XX
CC   R75669-R75678 and R75682 are streptokinase binding peptides derived
CC   from the human plasmin B chain in the regions 576-600, 620-640 or
CC   640-655 amino acids. The peptides are also capable of binding other
CC   proteins useful particularly when immobilised on a solid support for
CC   use in affinity chromatography.
XX
SQ   Sequence   7 AA;

  Query Match             42.2%;  Score 19;  DB 16;  Length 7;
  Best Local Similarity   75.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 NINL 9
            |:||
Db        1 nvnl 4


RESULT  30
W12026
ID   W12026 standard; peptide; 7 AA.
XX
AC   W12026;
XX
DT   06-NOV-1997  (first entry)
XX
DE   Herpes virus DNA polymerase RFHV specific antigen.
XX
KW   Retroperitoneal fibromatosis herpes virus; detection; infection;
KW   Kaposi's sarcoma herpes virus; viral DNA; viral RNA; vaccine;
KW   antigen; antibody.
XX
OS   Retroperitoneal fibromatosis herpes virus.
XX
PN   WO9704105-A1.
XX
PD   06-FEB-1997.
XX
PF   12-JUL-1996;   96WO-US11688.
XX
PR   11-JUL-1996;   96US-0001148.
PR   14-JUL-1995;   95US-0001148.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Bosch ML,  Rose TM,  Strand K,  Todaro GJ;
XX
DR   WPI; 1997-132644/12.
XX
PT   Herpes virus DNA polymerase and corresponding nucleotide sequence -
```

```
PT   used in the detection and treatment of herpes virus infection
XX
PS   Claim 15; Page 45; 132pp; English.
XX
CC   The present sequence represents a retroperitoneal fibromatosis herpes
CC   virus (RFHV) DNA polymerase specific antigen. Primers used in the
CC   amplification of polynucleotides encoding DNA polymerases may be used
CC   for detecting viral DNA or RNA in a sample of primate origin,
CC   especially in the diagnosis of herpes viral infection. Herpes virus
CC   DNA polymerases of this invention, may be used in vaccines for the
CC   protection against infection by a herpes virus of the RFHV/KSHV family.
CC   They may also be used in the design and screening of anti-viral drugs.
CC   Antibodies raised against the polymerase or fragments of it, may be
CC   used in the detection of herpes virus infection and for drug targetting
CC   for the therapy of herpes virus infection.
XX
SQ   Sequence   7 AA;

  Query Match             42.2%;  Score 19;  DB 18;  Length 7;
  Best Local Similarity   42.9%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSGANI 7
             :|||  :
Db        1 hlsggtv 7


RESULT  31
R34629
ID   R34629 standard; Protein; 8 AA.
XX
AC   R34629;
XX
DT   05-AUG-1993  (first entry)
XX
DE   B. thuringiensis dipteran toxin probe (m) design peptide.
XX
KW   Probe; Bacillus thuringiensis; Bt; insecticidal; active; endotoxin;
KW   lepidoptera; diptera; coleoptera.
XX
OS   Synthetic.
XX
PN   US5204237-A.
XX
PD   20-APR-1993.
XX
PF   25-OCT-1989;   89US-0427068.
XX
PR   25-OCT-1989;   89US-0427068.
PR   26-JUL-1991;   91US-0737569.
XX
PA   (MYCO ) MYCOGEN CORP.
XX
PI   Gaertner FH,  Schwab GE,  Sick AJ;
XX
DR   WPI; 1993-143952/17.
XX
PT   Gene probe for identifying endotoxin sequences in Bacillus
PT   thuringiensis - are universal, type specific or gene specific,
PT   for rapid detection and characterisation of insecticidal activity
XX
PS   Disclosure; Column 2; 5pp; English.
XX
CC   The sequences given in R35090-100 and R34628-32 are peptides which
CC   were used to design the probes given in Q40219-34.  These probes can
CC   be used to identify Bacillus thuringiensis (Bt) DNA which encodes
CC   insecticidally active endotoxin.  Probe (a) identifies genes encoding
CC   a toxin effective against any sort of insect.  Sequences (b) and (e)-
CC   (l) are specific for toxins against lepidoptera; sequences (c), (m)
CC   and (n) for toxins active against diptera and sequence (d), (o) and
CC   (p) for toxins active against coleoptera.
XX
```

```
SQ   Sequence   8 AA;

  Query Match             42.2%;  Score 19;  DB 14;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GANIN 8
           | |:|
Db       2 gtnmn 6


RESULT  32
R78218
ID   R78218 standard; Peptide; 8 AA.
XX
AC   R78218;
XX
DT   22-FEB-1996  (first entry)
XX
DE   B. thuringiensis endotoxin derived peptide.
XX
KW   Endotoxin; probe; microbes; peptide; detection; dipteran.
XX
OS   Bacillus thuringiensis.
XX
PN   US5430137-A.
XX
PD   04-JUL-1995.
XX
PF   25-OCT-1989;   89US-0427068.
XX
PR   30-OCT-1992;   92US-0968781.
PR   25-OCT-1989;   89US-0427068.
PR   26-JUL-1991;   91US-0737569.
XX
PA   (MYCO ) MYCOGEN CORP.
XX
PI   Gaertner FH,  Narva KE,  Schnepf HE,  Schwab GE,  Sick AJ;
PI   Thompson M;
XX
DR   WPI; 1995-245777/32.
XX
PT   Nucleotide sequence used as probes to identify Bacillus
PT   thuringiensis - are derived from the B.thuringiensis endotoxin
PT   genes, for identifying microbes which encode toxins
XX
PS   Claim 1; Columns 49-50; 30pp; English.
XX
CC   The nucleotide sequences which encode R78218 and R78219, B.
CC   thuringiensis (B.t.) endotoxin derived peptides, specify the probes
CC   Q94867 and Q94868 respectively. The probes can be used for the
CC   detection of endotoxin producing B.t. microbes. The probes aid in
CC   the search for useful microbes hosting toxin encoding genes,
CC   specifically from dipteran species.
XX
SQ   Sequence   8 AA;

  Query Match             42.2%;  Score 19;  DB 16;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GANIN 8
           | |:|
Db       2 gtnmn 6


RESULT  33
R23738
ID   R23738 standard; Protein; 9 AA.
XX
AC   R23738;
XX
DT   27-OCT-1992  (first entry)
XX
DE   Ligand CDR(f).
XX
KW   Complementarity determining region; heavy chain variable domain;
KW   antigen binding site.
XX
OS   Synthetic.
XX
PN   WO9206193-A.
XX
PD   16-APR-1992.
XX
PF   04-OCT-1991;   91WO-GB01726.
XX
PR   05-OCT-1990;   90GB-0021679.
XX
PA   (GORM/) GORMAN S D.
XX
PI   Gorman SD,  Routledge EG,  Waldmann H;
XX
DR   WPI; 1992-150879/18.
DR   N-PSDB; Q24342.
XX
PT   Ligands and antibodies with binding affinity for CD3 antigen -
PT   for treatment of immunosuppression e.g. in graft rejection, and
PT   cancer, esp. lymphoid malignancies
XX
PS   Disclosure; Page 4; 49pp; English.
XX
CC   The sequence given is a complementarity determining region (CDR)
CC   designated CDR(f).  CDR's are found in the variable domains of light
CC   and heavy chains which form the antigen binding site, and act as
CC   connectors between the four framework regions.
CC   It has been noted that there seem to be no characteristic features
CC   which distinguish human from mouse or rat CDR's and they are
CC   therefore immunologically identical.  The CDR described is used in
CC   a ligand with a binding affinity for the human CD3 antigen within
CC   the framework regions of the heavy chain variable domain along with
CC   the CDR's decribed in R23736 and R23737.
XX
SQ   Sequence   9 AA;

  Query Match             42.2%;  Score 19;  DB 13;  Length 9;
  Best Local Similarity   42.9%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGANI 7
           |:|  |:
Db       3 yvssfnv 9


RESULT  34
R37221
ID   R37221 standard; peptide; 9 AA.
XX
AC   R37221;
XX
DT   06-SEP-1993  (first entry)
XX
DE   IL-6 antagonist peptide #5.
XX
KW   IL-6; antagonist; cytokine; B cell differentiation; inflammation;
KW   tissue injury; B9.9 hybridoma cell line; Castleman's Disease;
KW   Lennart's T-cell lymphoma; Non-Hodgkin's lymphoma; allergy;
KW   immune deficiency disease; cardiac myxoma;
KW   mesangial proliferative glomerulonephritis.
XX
PN   US5210075-A.
XX
PD   11-MAY-1993.
```

```
XX
PF   16-FEB-1990;   90US-0480868.
XX
PR   16-FEB-1990;   90US-0480868.
XX
PA   (TANA ) TANABE SEIYAKU CO.
XX
PI   Chiang S,  Lobl TJ,  Nagarajan G,  Scholz W;
XX
DR   WPI; 1993-166970/20.
XX
PT   New peptide interleukin-6 antagonists - for treating and
PT   preventing auto-immune, immuno:inflammatory, neoplastic and
PT   infectious diseases etc.
XX
PS   Example 3; Column 27; 20pp; English.
XX
CC   This peptide is a specific example of a highly generic claimed
CC   formula covering sequences derived from the p51-70 portion (i.e the
CC   IL-6 receptor-binding portion) of IL-6 (see R37216) or modelled
CC   after different portions of this sequence. In an assay to determine
CC   IL-6 antagonist activity of this peptide, proliferation of the IL-6
CC   dependent B9.9 hybridoma cell line (J.Immunol. 139: 4116, 1987) was
CC   inhibited.  See R37216-R37261.
XX
SQ   Sequence   9 AA;


  Query Match             42.2%;  Score 19;  DB 14;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        2 LSGANINL 9
            |:  |:||
Db        1 laennlnl 8


RESULT  35
R50110
ID   R50110 standard; peptide; 9 AA.
XX
AC   R50110;
XX
DT   27-APR-1994  (first entry)
XX
DE   CDR (f).
XX
KW   Antibody; glycosylation; IgG; binding affinity; CD3; antigen;
KW   CDR; complementarity determining region; heavy chain; light chain;
KW   variable domain; immunosuppressant; T-cell; mitogenesis;
KW   cytokine; graft; rejection; cancer; proliferation.
XX
OS   Synthetic.
XX
PN   WO9319196-A.
XX
PD   30-SEP-1993.
XX
PF   21-OCT-1992;   92WO-GB01933.
XX
PR   24-MAR-1992;   92GB-0006422.
XX
PA   (BOLT/) BOLT S L.
PA   (CLAR/) CLARK M R.
PA   (GORM/) GORMAN S D.
PA   (ROUT/) ROUTLEDGE E G.
PA   (WALD/) WALDMANN H.
XX
PI   Bolt SL,  Clark MR,  Gorman SD,  Routledge EG,  Waldmann H;
XX
DR   WPI; 1993-320762/40.
XX
PT   Aglycosylated IgG antibody against human CD3 antigen - for
PT   preventing graft rejection and treating cancer, and does not
PT   induce T cell proliferation and cytokine prodn.
XX
PS   Claim 3; Page 25; 41pp; English.
XX
CC   An aglycosylated IgG antibody having a binding affinity for the human
CC   CD3 antigen complex comprises at least one CDR selected from the amino
CC   acid sequences given in R50105-R50110.  The heavy and light chain
CC   variable domains of an aglycosylated antibody is given in R41721-
CC   R41722.  The antibodies are useful as immunosupressants.  Unlike
CC   glycosylated antibodies, they do not induce T-cell mitogenesis nor
CC   do they cause high level release of cytokines.
XX
SQ   Sequence   9 AA;


  Query Match             42.2%;  Score 19;  DB 14;  Length 9;
  Best Local Similarity   42.9%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSGANI 7
            |:|  |:
Db        3 yvssfnv 9


Search completed: December 16, 2000, 03:07:19
Job time: 8060 sec
```

```
                   GenCore version 4.5
         Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 02:00:57 ; Search time 113.2 Seconds
                                              (without alignments)
                                              7.424 Million cell updates/sec

Title:          US-09-529-121-3
Perfect score:  45
Sequence:       1 YLSGADINL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters:      297973

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      SPTREMBL_14:*
               1:  sp_archea:*
               2:  sp_bacteria:*
               3:  sp_fungi:*
               4:  sp_human:*
               5:  sp_invertebrate:*
               6:  sp_mammal:*
               7:  sp_mhc:*
               8:  sp_organelle:*
               9:  sp_phage:*
               10:  sp_plant:*
               11:  sp_rodent:*
               12:  sp_virus:*
               13:  sp_vertebrate:*
               14:  sp_unclassified:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                 %
Result          Query
  No.   Score   Match Length DB  ID                          Description
---------------------------------------------------------------------------
    1      35    77.8    352 12  Q65258                       Q65258 african swi
    2      34    75.6    176  1  O58971                      O58971 pyrococcus
    3      34    75.6    763  4  Q16679                      Q16679 homo sapien
    4      34    75.6    860  2  Q9WZL6                      Q9wzl6 thermotoga
    5      33    73.3    171  4  Q99521                      Q99521 homo sapien
    6      33    73.3    401  5  O02314                      O02314 caenorhabdi
    7      33    73.3    402  5  O02315                      O02315 caenorhabdi
    8      33    73.3    441  5  Q19407                      Q19407 caenorhabdi
    9      33    73.3    445  5  Q19730                      Q19730 caenorhabdi
   10      33    73.3    454  5  Q19729                      Q19729 caenorhabdi
   11      33    73.3    519  2  Q47916                      Q47916 fibrobacter
   12      33    73.3    663  4  Q9ULC6                      Q9ulc6 homo sapien
   13      33    73.3    705  2  O31988                      O31988 bacillus su
   14      33    73.3    705  9  O64034                      O64034 bacteriopha
   15      33    73.3    707  2  O30671                      O30671 bacillus su
   16      33    73.3    880  3  Q06708                      Q06708 saccharomyc
   17      33    73.3   1006 10  Q9SA76                       Q9sa76 arabidopsis
   18      32    71.1    188  2  O67835                      O67835 aquifex aeo
   19      32    71.1    307 10  Q9SDD1                       Q9sdd1 oryza sativ
   20      32    71.1    308  5  Q9V3H2                      Q9v3h2 drosophila
   21      32    71.1    309 11  O35593                       O35593 mus musculu
   22      32    71.1    310  4  O00487                      O00487 homo sapien
   23      32    71.1    312  5  O76577                      O76577 caenorhabdi
   24      32    71.1    350  5  Q20618                      Q20618 caenorhabdi
   25      32    71.1    352  3  Q9Y887                      Q9y887 suillus bov
   26      32    71.1    374  1  O27425                      O27425 methanobact
   27      32    71.1    413  5  Q9V405                      Q9v405 drosophila
   28      32    71.1    418  4  Q9UEX3                      Q9uex3 homo sapien
   29      32    71.1    418  4  Q9UBM3                      Q9ubm3 homo sapien
   30      32    71.1    432  3  Q9UW41                      Q9uw41 cryptococcu
   31      32    71.1    540  2  O31716                      O31716 bacillus su
   32      32    71.1    635  2  P94146                      P94146 alcaligenes
   33      32    71.1    763  6  O62831                      O62831 bos taurus
   34      32    71.1    773  6  O62828                      O62828 bos taurus
   35      32    71.1   1650 11  Q9QVT6                       Q9qvt6 rattus sp.
   36      32    71.1   1924  4  Q9UPY3                      Q9upy3 homo sapien
   37.     31    68.9    131  8  Q35903                      Q35903 strongyloce
   38      31    68.9    175  1  Q9UZT7                      Q9uzt7 pyrococcus
   39      31    68.9    263  5  Q9VGM7                      Q9vgm7 drosophila
   40      31    68.9    292  5  Q9VGN0                      Q9vgn0 drosophila
   41      31    68.9    301  2  Q9RW36                      Q9rw36 deinococcus
   42      31    68.9    404 10  Q9SYB9                       Q9syb9 arabidopsis
   43      31    68.9    409  2  O24843                      O24843 acinetobact
   44      31    68.9    459 10  Q9SS04                       Q9ss04 arabidopsis
   45      31    68.9    461  6  O97884                      O97884 equus cabal

                                   ALIGNMENTS


RESULT   1
Q65258
ID   Q65258      PRELIMINARY;      PRT;   352 AA.
AC   Q65258;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   ORF L3R.
OS   African swine fever virus (ASFV).
OC   Viruses; dsDNA viruses, no RNA stage;
OC   African swine fever-like viruses.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MALAWI LIL20 /1;
RX   MEDLINE; 94014996.
RA   Vydelingum S., Baylis S.A., Bristow C., Smith G.L., Dixon L.K.;
RT   "Duplicated genes within the variable right end of the genome of a
RT   pathogenic isolate of African swine fever virus.";
RL   J. Gen. Virol. 74:2125-2130(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MALAWI LIL20 /1;
RX   MEDLINE; 94292916.
RA   Dixon L.K., Twigg S.R.F., Baylis S.A., Vydelingum S., Bristow C.,
RA   Hammond J.M., Smith G.L.;
RT   "Nucleotide sequence of a 55 kbp region from the right end of the
RT   genome of a pathogenic African swine fever virus isolate (Malawi
RT   LIL20/1).";
RL   J. Gen. Virol. 7:1655-1684(1994).
DR   EMBL; X71982; CAA50855.1; -.
DR   INTERPRO; IPR002595; -.
DR   PFAM; PF01671; ASFV_360; 1.
DR   PRODOM; PD003462; -; 1.
SQ   SEQUENCE   352 AA;  40682 MW;  61561D08AE1C1599 CRC64;


  Query Match             77.8%;  Score 35;  DB 12;  Length 352;
  Best Local Similarity   87.5%;  Pred. No. 25;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGADIN 8
           || |||||
```

```
Db    258 YLLGADIN 265


RESULT   2
O58971
ID   O58971      PRELIMINARY;      PRT;   176 AA.
AC   O58971;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   176AA LONG HYPOTHETICAL THERMONUCLEASE.
GN   PH1212.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE; 98344137.
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000005; BAA30312.1; -.
DR   HSSP; P00644; 1SNQ.
DR   INTERPRO; IPR000080; -.
DR   INTERPRO; IPR002071; -.
DR   PFAM; PF00565; SNase; 1.
DR   PROSITE; PS01123; TNASE_1; 1.
DR   PRODOM; PD002274; -; 1.
SQ   SEQUENCE   176 AA;  20010 MW;  CB802A467B17E29E CRC64;


  Query Match             75.6%;  Score 34;  DB 1;  Length 176;
  Best Local Similarity   75.0%;  Pred. No. 18;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGADIN 8
          ||:| |||
Db    128 YLNGTDIN 135


RESULT   3
Q16679
ID   Q16679      PRELIMINARY;      PRT;   763 AA.
AC   Q16679;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   GASTRIN-BINDING PROTEIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95002180.
RA   Zhang Q.X., Baldwin G.S.;
RT   "Structures of the human cDNA and gene encoding the 78 kDa gastrin-
RT   binding protein and of a related pseudogene.";
RL   Biochim. Biophys. Acta 1219:567-575(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Baldwin G.S.;
RL   Comp. Biochem. Physiol. 104:55-61(1993).
DR   EMBL; U04627; AAA56664.1; -.
DR   HSSP; P14604; 2DUB.
DR   INTERPRO; IPR001128; -.
DR   INTERPRO; IPR001753; -.
DR   INTERPRO; IPR002135; -.
DR   PFAM; PF00378; ECH; 1.
DR   PFAM; PF00725; 3HCDH; 1.
DR   PROSITE; PS00067; 3HCDH; 1.
DR   PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
DR   PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
SQ   SEQUENCE   763 AA;  83041 MW;  CDFB85DDE3E512E4 CRC64;


  Query Match             75.6%;  Score 34;  DB 4;  Length 763;
  Best Local Similarity   55.6%;  Pred. No. 1e+02;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YLSGADINL 9
          :::|||||:
Db     98 FIAGADINM 106


RESULT   4
Q9WZL6
ID   Q9WZL6      PRELIMINARY;      PRT;   860 AA.
AC   Q9WZL6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HYPOTHETICAL 101.0 KDA PROTEIN.
GN   TM0757.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogales; Thermotoga.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE; 99287316.
RA   Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA   Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA   McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA   Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA   Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA   Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT   "Evidence for lateral gene transfer between Archaea and Bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
DR   EMBL; AE001745; AAD35839.1; -.
DR   TIGR; TM0757; -.
DR   INTERPRO; IPR001173; -.
DR   INTERPRO; IPR001440; -.
DR   PFAM; PF00515; TPR; 1.
DR   PFAM; PF00535; Glycos_transf_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   860 AA;  100968 MW;  7781CCFA62FB89CF CRC64;


  Query Match             75.6%;  Score 34;  DB 2;  Length 860;
  Best Local Similarity   75.0%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGADIN 8
          ||:| |||
Db    443 YLTGGDIN 450


RESULT   5
Q99521
ID   Q99521      PRELIMINARY;      PRT;   171 AA.
AC   Q99521;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   GASTRIN BINDING PROTEIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE; 97352549.
RA   Monstein H.J., Nylander A.G., Hakanson R.;
RT   "Widespread tissue expression of gastrin-binding-protein mRNA.";
RL   Eur. J. Biochem. 246:502-507(1997).
DR   EMBL; X98225; CAA66885.1; -.
DR   HSSP; P14604; 2DUB.
DR   INTERPRO; IPR001753; -.
DR   PFAM; PF00378; ECH; 1.
DR   PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
FT   NON_TER       1      1
FT   NON_TER     171    171
SQ   SEQUENCE   171 AA;  18141 MW;  7CC22F2E66B98447 CRC64;

  Query Match             73.3%;  Score 33;  DB 4;  Length 171;
  Best Local Similarity   55.6%;  Pred. No. 28;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADINL 9
            :::|||||:
Db       65 FVAGADINM 73


RESULT   6
O02314
ID   O02314      PRELIMINARY;      PRT;   401 AA.
AC   O02314;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   T15D6.2 PROTEIN.
GN   T15D6.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Dobson R.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z83125; CAB05620.1; -.
SQ   SEQUENCE   401 AA;  45608 MW;  E1D3F8AD84C0873D CRC64;

  Query Match             73.3%;  Score 33;  DB 5;  Length 401;
  Best Local Similarity   77.8%;  Pred. No. 78;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSGADINL 9
            |||| || |
Db      186 YLSGVDIPL 194


RESULT   7
O02315
ID   O02315      PRELIMINARY;      PRT;   402 AA.
AC   O02315;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   T15D6.3 PROTEIN.
GN   T15D6.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Dobson R.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z83125; CAB05621.1; -.
SQ   SEQUENCE   402 AA;  46135 MW;  F6AD721976703307 CRC64;

  Query Match             73.3%;  Score 33;  DB 5;  Length 402;
  Best Local Similarity   77.8%;  Pred. No. 78;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSGADINL 9
            |||| || |
Db      196 YLSGVDIPL 204


RESULT   8
Q19407
ID   Q19407      PRELIMINARY;      PRT;   441 AA.
AC   Q19407;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   F13E9.1 PROTEIN.
GN   F13E9.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mcmurray A.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z69383; CAA93405.1; -.
```

```
DR   INTERPRO; IPR001611; -.
DR   PFAM; PF00560; LRR; 4.
SQ   SEQUENCE   441 AA;  50845 MW;  66AC257B7884CCFA CRC64;

  Query Match             73.3%;  Score 33;  DB 5;  Length 441;
  Best Local Similarity   62.5%;  Pred. No. 87;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGADIN 8
           |:|| |:|
Db     427 YISGEDVN 434


RESULT   9
Q19730
ID   Q19730      PRELIMINARY;      PRT;   445 AA.
AC   Q19730;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   F22D6.11 PROTEIN.
GN   F22D6.11.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wilkinson J.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z71262; CAA95817.1; -.
SQ   SEQUENCE   445 AA;  51330 MW;  4380A86BAF05DC6D CRC64;

  Query Match             73.3%;  Score 33;  DB 5;  Length 445;
  Best Local Similarity   77.8%;  Pred. No. 88;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
           |||| || |
Db     215 YLSGVDIPL 223


RESULT  10
Q19729
ID   Q19729      PRELIMINARY;      PRT;   454 AA.
AC   Q19729;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   F22D6.12 PROTEIN.
GN   F22D6.12.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wilkinson J.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z71262; CAA95816.1; -.
SQ   SEQUENCE   454 AA;  52822 MW;  FA0134DBF31D0925 CRC64;

  Query Match             73.3%;  Score 33;  DB 5;  Length 454;
  Best Local Similarity   77.8%;  Pred. No. 90;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
           |||| || |
Db     225 YLSGVDIPL 233


RESULT  11
Q47916
ID   Q47916      PRELIMINARY;      PRT;   519 AA.
AC   Q47916;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ENDOGLUCANASE CELG.
OS   Fibrobacter succinogenes (Bacteroides succinogenes).
OC   Bacteria; Fibrobacteria; Fibrobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S85;
RX   MEDLINE; 97017599.
RA   Iyo A.H., Forsberg C.W.;
RT   "Endoglucanase G from Fibrobacter succinogenes S85 belongs to a class
RT   of enzymes characterized by a basic C-terminal domain.";
RL   Can. J. Microbiol. 42:934-943(1996).
DR   EMBL; U33887; AAB38548.1; -.
DR   HSSP; P17901; 1EDG.
DR   INTERPRO; IPR001547; -.
DR   PFAM; PF00150; cellulase; 1.
SQ   SEQUENCE   519 AA;  56848 MW;  B06D2113B10FF27E CRC64;

  Query Match             73.3%;  Score 33;  DB 2;  Length 519;
  Best Local Similarity   75.0%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGADINL 9
           |||:|:||
Db     334 LSGSDLNL 341


RESULT  12
Q9ULC6
ID   Q9ULC6      PRELIMINARY;      PRT;   663 AA.
AC   Q9ULC6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   PEPTIDYLARGININE DEIMINASE TYPE I (EC 3.5.3.15).
```

```
GN   HPAD-COLONY10.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EPIDERMIS;
RA   Ishigami A., Guerrin-Weber M., Sebbag M., Serre G., Senshu T.;
RT   "Molecular cloning of peptidylarginine deiminase type I cDNA from a
RT   human epidermis.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB033768; BAA85771.1; -.
KW   Hydrolase.
SQ   SEQUENCE   663 AA;  74607 MW;  0BD8D460634EE2D9 CRC64;

  Query Match             73.3%;  Score 33;  DB 4;  Length 663;
  Best Local Similarity   66.7%;  Pred. No. 1.4e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGADINL 9
            ||:| ||:|
Db      111 YLTGVDISL 119

RESULT  13
O31988
ID   O31988      PRELIMINARY;      PRT;   705 AA.
AC   O31988;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   SUBLANCIN 168 LANTIBIOTIC TRANSPORTER.
GN   SUNT.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE; 98044033.
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z99115; CAB14065.1; -.
DR   INTERPRO; IPR001140; -.
DR   INTERPRO; IPR001617; -.
DR   INTERPRO; IPR002078; -.
DR   INTERPRO; IPR002106; -.
DR   PFAM; PF00005; ABC_tran; 1.
DR   PFAM; PF00664; ABC_membrane; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.
SQ   SEQUENCE   705 AA;  81563 MW;  0F94D9964AF8BD3A CRC64;

  Query Match             73.3%;  Score 33;  DB 2;  Length 705;
  Best Local Similarity   75.0%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGADIN 8
            ||:| |||
Db      541 YLNGLDIN 548

RESULT  14
O64034
ID   O64034      PRELIMINARY;      PRT;   705 AA.
AC   O64034;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   ABC TRANSPORTER.
GN   YOLH.
OS   Bacteriophage SPBc2.
OC   Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Mauel C.,
RA   Karamata D.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF020713; AAC12993.1; -.
DR   INTERPRO; IPR001140; -.
DR   INTERPRO; IPR001617; -.
DR   INTERPRO; IPR002078; -.
DR   INTERPRO; IPR002106; -.
DR   PFAM; PF00005; ABC_tran; 1.
DR   PFAM; PF00664; ABC_membrane; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.
SQ   SEQUENCE   705 AA;  81563 MW;  0F94D9964AF8BD3A CRC64;

  Query Match             73.3%;  Score 33;  DB 9;  Length 705;
  Best Local Similarity   75.0%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGADIN 8
            ||:| |||
Db      541 YLNGLDIN 548

RESULT  15
O30671
ID   O30671      PRELIMINARY;      PRT;   707 AA.
AC   O30671;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   SUBLANCIN TRANSPORTER PROTEIN.
GN   SUNT.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
```

```
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE; 98389743.
RA   Paik S.H., Chakicherla A., Hansen J.N.;
RT   "Identification and characterization of the structural and transporter
RT   genes for, and the chemical and biological properties of, sublancin
RT   168, a novel lantibiotic produced by Bacillus subtilis 168.";
RL   J. Biol. Chem. 273:23134-23142(1998).
DR   EMBL; AF014938; AAC63532.1; -.
DR   INTERPRO; IPR001140; -.
DR   INTERPRO; IPR001617; -.
DR   INTERPRO; IPR002078; -.
DR   INTERPRO; IPR002106; -.
DR   PFAM; PF00005; ABC_tran; 1.
DR   PFAM; PF00664; ABC_membrane; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.
SQ   SEQUENCE   707 AA;  81805 MW;  ACF7F9083B00422C CRC64;


  Query Match             73.3%;  Score 33;  DB 2;  Length 707;
  Best Local Similarity   75.0%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGADIN 8
           ||:| |||
Db     543 YLNGLDIN 550


Search completed: December 16, 2000, 02:00:57
Job time: 5931 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 03:05:19 ; Search time 69.45 Seconds
                                              (without alignments)
                                              4.140 Million cell updates/sec

Title:           US-09-529-121-3
Perfect score:   45
Sequence:        1 YLSGADINL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters:      87993

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_39:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                 %
Result           Query
  No.   Score   Match Length DB  ID                             Description
-------------------------------------------------------------------------------
    1      38    84.4    702  1  CCEM_HUMAN                   P06731 homo sapien
    2      36    80.0    455  1  PEX3_PICPA                   Q92262 pichia past
    3      34    75.6    572  1  GPC5_HUMAN                   P78333 homo sapien
    4      34    75.6    714  1  YFCX_ECOLI                   P77399 escherichia
    5      34    75.6    763  1  ECHA_HUMAN                   P40939 homo sapien
    6      34    75.6    763  1  ECHA_PIG                     Q29554 sus scrofa
    7      34    75.6   1986  1  WA_EMENI                     Q03149 emericella
    8      33    73.3    311  1  V311_ASFB7                   P23163 african swi
    9      33    73.3    763  1  ECHA_RAT                     Q64428 rattus norv
   10      33    73.3    880  1  YL86_YEAST                   Q06708 saccharomyc
   11      32    71.1    363  1  PGL1_COLLN                   Q00446 colletotric
   12      32    71.1    414  1  PRS6_CAEEL                   P46502 caenorhabdi
   13      32    71.1    418  1  PRS6_HUMAN                   P43686 homo sapien
   14      32    71.1    418  1  PRS6_MOUSE                   P54775 mus musculu
   15      32    71.1    418  1  PRS6_RAT                     Q63570 rattus norv
   16      32    71.1   4660  1  LRP2_RAT                     P98158 rattus norv
   17      31    68.9     30  1  RET1_BOVIN                   P02694 bos taurus
   18      31    68.9    108  1  HLYU_VIBCH                   P52695 vibrio chol
   19      31    68.9    172  1  PHEB_ANASP                   P35797 anabaena sp
   20      31    68.9    210  1  YRL1_METVA                   P15886 methanococc
   21      31    68.9    340  1  Y05A_BPT4                    P39256 bacteriopha
   22      31    68.9    366  1  FTSZ_PYRWO                   Q52630 pyrococcus
   23      31    68.9    415  1  PRS6_MANSE                   P46507 manduca sex
   24      31    68.9    491  1  PM17_BOVIN                   Q06154 bos taurus
   25      31    68.9    502  1  YEI0_YEAST                   P39941 saccharomyc
   26      31    68.9    580  1  GPC3_HUMAN                   P51654 homo sapien
   27      31    68.9    586  1  LRE1_YEAST                   P25579 saccharomyc
   28      31    68.9    597  1  GPC3_RAT                     P13265 rattus norv
   29      31    68.9    699  1  EFG_RICPR                    P41084 rickettsia
   30      31    68.9    854  1  LDLR_CRIGR                   P35950 cricetulus
   31      31    68.9    860  1  LDLR_HUMAN                   P01130 homo sapien
   32      31    68.9    864  1  LDLR_MOUSE                   P35951 mus musculu
   33      31    68.9    879  1  LDLR_RAT                     P35952 rattus norv
   34      31    68.9   2109  1  RRPL_VSVJH                   P13615 vesicular s
   35      31    68.9   4655  1  LRP2_HUMAN                   P98164 homo sapien
   36      30    66.7    117  1  V1P_MOUSE                    P80144 mus musculu
   37      30    66.7    118  1  V1P_CHICK                    Q91955 gallus gall
   38      30    66.7    167  1  LKC1_PASHA                   P16533 pasteurella
   39      30    66.7    167  1  LKC3_PASHA                   P55120 pasteurella
   40      30    66.7    167  1  LKCB_PASHA                   P55121 pasteurella
   41      30    66.7    177  1  IL1X_HUMAN                   P18510 homo sapien
   42      30    66.7    177  1  TRF4_ECOLI                   Q03450 escherichia
   43      30    66.7    184  1  RS13_SULAC                   P39470 sulfolobus
   44      30    66.7    248  1  PSPA_MOUSE                   P35242 mus musculu
   45      30    66.7    248  1  PSPA_RAT                     P08427 rattus norv

                                ALIGNMENTS

RESULT   1
CCEM_HUMAN
ID   CCEM_HUMAN     STANDARD;      PRT;   702 AA.
AC   P06731;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINOEMBRYONIC ANTIGEN PRECURSOR (CEA) (MECONIUM ANTIGEN 100)
DE   (CD66E ANTIGEN).
GN   CEA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90258861.
RA   Schrewe H., Thompson J., Bona M., Hefta L.J.F., Maruya A.,
RA   Hassauer M., Shively J.E., von Kleist S., Zimmermann W.;
RT   "Cloning of the complete gene for carcinoembryonic antigen: analysis
RT   of its promoter indicates a region conveying cell type-specific
RT   expression.";
RL   Mol. Cell. Biol. 10:2738-2748(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88038876.
RA   Beauchemin N., Benchimol S., Cournoyer D., Fuks A., Stanners C.P.;
RT   "Isolation and characterization of full-length functional cDNA clones
RT   for human carcinoembryonic antigen.";
RL   Mol. Cell. Biol. 7:3221-3230(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89122014.
RA   Barnett T., Goebel S.J., Nothdurft M.A., Elting J.J.;
RT   "Carcinoembryonic antigen family: characterization of cDNAs coding
RT   for NCA and CEA and suggestion of nonrandom sequence variation in
RT   their conserved loop-domains.";
RL   Genomics 3:59-66(1988).
RN   [4]
RP   SEQUENCE OF 5-702 FROM N.A.
RX   MEDLINE; 87128144.
RA   Oikawa S., Nakazato H., Kosaki G.;
RT   "Primary structure of human carcinoembryonic antigen (CEA) deduced
RT   from cDNA sequence.";
RL   Biochem. Biophys. Res. Commun. 142:511-518(1987).
RN   [5]
RP   SEQUENCE OF 331-702 FROM N.A.
RX   MEDLINE; 87204247.
RA   Zimmermann W., Ortlieb B., Friedrich R., von Kleist S.;
RT   "Isolation and characterization of cDNA clones encoding the human
RT   carcinoembryonic antigen reveal a highly conserved repeating
RT   structure.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:2960-2964(1987).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC   -!- TISSUE SPECIFICITY: FOUND IN ADENOCARCINOMAS OF ENDODERMALLY
CC       DERIVED DIGESTIVE SYSTEM EPITHELIUM AND FETAL COLON.
CC   -!- PTM: COMPLEX IMMUNOREACTIVE GLYCOPROTEIN WITH A MW OF 180 KDA
```

```
CC       COMPRISING 60% CARBOHYDRATE.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       7 IG-LIKE DOMAINS. BELONGS TO THE CARCINOEMBRYONIC ANTIGEN
CC       SUBFAMILY.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD66e entry;
CC       WWW="HTTP://WWW.NCBI.NLM.NIH.GOV/PROW/CD/CD66E.HTM".
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M17303; AAB59513.1; -.
DR   EMBL; M59262; AAA62835.1; ALT_SEQ.
DR   EMBL; M59255; AAA62835.1; JOINED.
DR   EMBL; M59256; AAA62835.1; JOINED.
DR   EMBL; M59257; AAA62835.1; JOINED.
DR   EMBL; M59258; AAA62835.1; JOINED.
DR   EMBL; M59259; AAA62835.1; JOINED.
DR   EMBL; M59260; AAA62835.1; JOINED.
DR   EMBL; M59261; AAA62835.1; JOINED.
DR   EMBL; M59709; -; NOT_ANNOTATED_CDS.
DR   EMBL; M59710; -; NOT_ANNOTATED_CDS.
DR   EMBL; M29540; AAA51967.1; -.
DR   EMBL; X16455; CAA34474.1; -.
DR   EMBL; M15042; AAA51963.1; -.
DR   EMBL; M16234; AAA51972.1; -.
DR   PIR; A36319; A36319.
DR   MIM; 114890; -.
DR   INTERPRO; IPR003006; -.
DR   PFAM; PF00047; ig; 7.
KW   Immunoglobulin domain; Glycoprotein; GPI-anchor; Membrane; Signal.
FT   SIGNAL        1     34
FT   CHAIN        35      ?       CARCINOEMBRYONIC ANTIGEN.
FT   PROPEP        ?    702       REMOVED IN MATURE FORM.
FT   DOMAIN       35    144       IG-LIKE DOMAIN 1.
FT   DOMAIN      146    237       IG-LIKE DOMAIN 2.
FT   DOMAIN      238    322       IG-LIKE DOMAIN 3.
FT   DOMAIN      324    415       IG-LIKE DOMAIN 4.
FT   DOMAIN      416    498       IG-LIKE DOMAIN 5.
FT   DOMAIN      502    593       IG-LIKE DOMAIN 6.
FT   DOMAIN      594    677       IG-LIKE DOMAIN 7.
FT   CARBOHYD    104    104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    115    115       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    152    152       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    182    182       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    197    197       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    204    204       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    208    208       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    246    246       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    256    256       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    274    274       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    288    288       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    292    292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    309    309       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    330    330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    351    351       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    360    360       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    375    375       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    432    432       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    466    466       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    480    480       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    508    508       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    529    529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    553    553       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    560    560       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    580    580       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    612    612       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    650    650       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    665    665       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CONFLICT    320    320       MISSING (IN REF. 4).
SQ   SEQUENCE   702 AA;  76795 MW;  6299AE26CDDBDB5C CRC64;


  Query Match             84.4%;  Score 38;  DB 1;  Length 702;
  Best Local Similarity   77.8%;  Pred. No. 3.1;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              |||||::||
Db        605 YLSGANLNL 613


RESULT   2
PEX3_PICPA
ID   PEX3_PICPA     STANDARD;      PRT;   455 AA.
AC   Q92262;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PEROXISOMAL MEMBRANE PROTEIN PAS2 (PEROXIN-3).
GN   PEX3 OR PAS2.
OS   Pichia pastoris (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Pichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97115764.
RA   Subramani S.;
RT   "Protein translocation into peroxisomes.";
RL   J. Biol. Chem. 271:32483-32486(1996).
CC   -!- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z72390; CAA96530.1; -.
KW   Transmembrane; Peroxisome.
FT   DOMAIN        1     15       MATRIX (POTENTIAL).
FT   TRANSMEM     16     33       POTENTIAL.
FT   DOMAIN       34    455       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   455 AA;  51973 MW;  6853C5EA5C67EC34 CRC64;


  Query Match             80.0%;  Score 36;  DB 1;  Length 455;
  Best Local Similarity   77.8%;  Pred. No. 5.1;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              ||: |||||
Db        353 YLNNADINL 361


RESULT   3
GPC5_HUMAN
ID   GPC5_HUMAN     STANDARD;      PRT;   572 AA.
AC   P78333; O60436;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLYPICAN-5 PRECURSOR.
GN   GPC5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   TISSUE=BRAIN;
RX   MEDLINE; 97224481.
RA   Veugelers M., Vermeesch J., Reekmans G., Steinfeld R., Marynen P.,
RA   David G.;
RT   "Characterization of glypican-5 and chromosomal localization of human
RT   GPC5, a new member of the glypican gene family.";
RL   Genomics 40:24-30(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 97470873.
RA   Saunders S., Paine-Saunders S., Lander A.D.;
RT   "Expression of the cell surface proteoglycan glypican-5 is
RT   developmentally regulated in kidney, limb, and brain.";
RL   Dev. Biol. 190:78-93(1997).
CC   -!- FUNCTION: HEPARAN SULFATE CELL SURFACE PROTEOGLYCAN (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC       SIMILARITY).
CC   -!- TISSUE SPECIFICITY: IN ADULT, PRIMARILY EXPRESSED IN THE BRAIN.
CC       ALSO DETECTED IN FETAL BRAIN, LUNG, AND LIVER.
CC   -!- SIMILARITY: BELONGS TO THE GLYPICAN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U66033; AAC51118.1; -.
DR   EMBL; AF001462; AAC12261.1; -.
DR   MIM; 602446; -.
DR   INTERPRO; IPR001863; -.
DR   PFAM; PF01153; Glypican; 1.
DR   PROSITE; PS01207; GLYPICAN; 1.
KW   Proteoglycan; Heparan sulfate; Glycoprotein; Signal; GPI-anchor.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25      ?       GLYPICAN-5.
FT   PROPEP        ?    572       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       13     18       POLY-LEU.
FT   CARBOHYD    120    120       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    237    237       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    527    527       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    441    441       O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT   CARBOHYD    486    486       O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT   CARBOHYD    495    495       O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT   CARBOHYD    507    507       O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT   CARBOHYD    509    509       O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT   CONFLICT    205    205       G -> C (IN REF. 2).
FT   CONFLICT    245    245       S -> F (IN REF. 2).
SQ   SEQUENCE   572 AA;  63707 MW;  A17969FE0DD0D308 CRC64;

  Query Match             75.6%;  Score 34;  DB 1;  Length 572;
  Best Local Similarity   75.0%;  Pred. No. 17;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGADIN 8
           || |||:|
Db     151 YLFGADVN 158


RESULT   4
YFCX_ECOLI
ID   YFCX_ECOLI     STANDARD;      PRT;   714 AA.
AC   P77399;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   PUTATIVE FATTY OXIDATION COMPLEX ALPHA SUBUNIT [INCLUDES: ENOYL-COA
DE   HYDRATASE (EC 4.2.1.17); 3-HYDROXYACYL-COA DEHYDROGENASE
DE   (EC 1.1.1.35); 3-HYDROXYBUTYRYL-COA EPIMERASE (EC 5.1.2.3)].
GN   YFCX.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 97349980.
RA   Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA   Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA   Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA   Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,
RA   Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA   Yamagata S., Horiuchi T.;
RT   "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT   - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT   analysis of its sequence features.";
RL   DNA Res. 4:91-113(1997).
CC   -!- CATALYTIC ACTIVITY: (3S)-3-HYDROXYACYL-COA = TRANS-2(OR 3)-ENOYL-
CC       COA + H(2)O.
CC   -!- CATALYTIC ACTIVITY: L-3-HYDROXYACYL-COA + NAD(+) = 3-OXOACYL-COA +
CC       NADH.
CC   -!- CATALYTIC ACTIVITY: (S)-3-HYDROXYBUTANOYL-COA =
CC       (R)-3-HYDROXYBUTANOYL-COA.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE ENOYL-COA
CC       HYDRATASE/ISOMERASE FAMILY.
CC   -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE 3-
CC       HYDROXYACYL-COA DEHYDROGENASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE000322; AAC75401.1; -.
DR   EMBL; D90864; CAB22124.1; -.
DR   EMBL; D90865; CAB22130.1; -.
DR   HSSP; P14604; 2DUB.
DR   ECOGENE; EG14127; YFCX.
DR   INTERPRO; IPR001753; -.
DR   INTERPRO; IPR002135; -.
DR   PFAM; PF00725; 3HCDH; 1.
DR   PFAM; PF00378; ECH; 1.
DR   PROSITE; PS00067; 3HCDH; 1.
DR   PROSITE; PS00166; ENOYL_COA_HYDRATASE; FALSE_NEG.
KW   Hypothetical protein; Fatty acid metabolism; Multifunctional enzyme;
KW   Oxidoreductase; NAD; Lyase; Isomerase.
FT   ACT_SITE    118    118       ACTIVATES A WATER MOLECULE (BY
FT                                SIMILARITY).
FT   ACT_SITE    140    140       PROVIDES THE ALPHA-PROTON (BY
FT                                SIMILARITY).
SQ   SEQUENCE   714 AA;  77072 MW;  F4E0A75680039A0D CRC64;

  Query Match             75.6%;  Score 34;  DB 1;  Length 714;
  Best Local Similarity   55.6%;  Pred. No. 22;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
```

```
         :::|||||:
Db     65 FIAGADINM 73


RESULT   5
ECHA_HUMAN
ID   ECHA_HUMAN     STANDARD;      PRT;   763 AA.
AC   P40939;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   MITOCHONDRIAL TRIFUNCTIONAL ENZYME ALPHA SUBUNIT PRECURSOR (TP-ALPHA)
DE   [INCLUDES: LONG-CHAIN ENOYL-COA HYDRATASE (EC 4.2.1.17); LONG CHAIN
DE   3-HYDROXYACYL-COA DEHYDROGENASE (EC 1.1.1.35)].
GN   HADHA OR HADH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94183263.
RA   Kamijo T., Aoyama T., Miyazaki J.I., Hashimoto T.;
RT   "Structural analysis of cDNAs for subunits of human mitochondrial
RT   fatty acid beta-oxidation trifunctional protein.";
RL   Biochem. Biophys. Res. Commun. 199:818-825(1994).
RN   [2]
RP   VARIANT AFLP GLN-510.
RX   MEDLINE; 95148633.
RA   Sims H.F., Brackett J.C., Powell C.K., Treem W.R., Hale D.E.,
RA   Bennett M.J., Gibson B., Shapiro S., Strauss A.W.;
RT   "The molecular basis of pediatric long chain 3-hydroxyacyl-CoA
RT   dehydrogenase deficiency associated with maternal acute fatty liver
RT   of pregnancy.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:841-845(1995).
CC   -!- FUNCTION: BIFUNCTIONAL SUBUNIT.
CC   -!- CATALYTIC ACTIVITY: (3S)-3-HYDROXYACYL-COA = TRANS-2(OR 3)-ENOYL-
CC       COA + H(2)O.
CC   -!- CATALYTIC ACTIVITY: L-3-HYDROXYACYL-COA + NAD(+) = 3-OXOACYL-COA +
CC       NADH.
CC   -!- PATHWAY: CATALYZES TWO STEPS OF THE FATTY ACID BETA-OXIDATION
CC       CYCLE.
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC   -!- DISEASE: DEFECTS IN HADHA ARE A CAUSE OF MATERNAL ACUTE FATTY
CC       LIVER OF PREGNANCY (AFLP), A SEVERE MATERNAL ILLNESS OCCURRING
CC       DURING PREGNANCIES WITH AFFECTED FETUSES. THIS DISEASE IS
CC       CHARACTERIZED BY SUDDEN UNEXPLAINED INFANT DEATH OR HYPOGLYCEMIA
CC       AND ABNORMAL LIVER ENZYMES (REYE-LIKE SYNDROME).
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE ENOYL-COA
CC       HYDRATASE/ISOMERASE FAMILY.
CC   -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE 3-
CC       HYDROXYACYL-COA DEHYDROGENASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as  long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D16480; BAA03941.1; -.
DR   PIR; JC2108; JC2108.
DR   HSSP; P14604; 2DUB.
DR   MIM; 600890; -.
DR   INTERPRO; IPR001753; -.
DR   INTERPRO; IPR002135; -.
DR   PFAM; PF00725; 3HCDH; 1.
DR   PFAM; PF00378; ECH; 1.
DR   PROSITE; PS00067; 3HCDH; 1.
DR   PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
KW   Fatty acid metabolism; Multifunctional enzyme; Oxidoreductase; NAD;
KW   Lyase; Mitochondrion; Transit peptide; Disease mutation.
FT   TRANSIT       1     36       MITOCHONDRION (POTENTIAL).
FT   CHAIN        37    763       MITOCHONDRIAL TRIFUNCTIONAL ENZYME ALPHA
FT                                SUBUNIT.
FT   ACT_SITE    151    151       ACTIVATES A WATER MOLECULE (BY
FT                                SIMILARITY).
FT   ACT_SITE    173    173       PROVIDES THE ALPHA-PROTON (BY
FT                                SIMILARITY).
FT   VARIANT     510    510       E -> Q (IN AFLP).
FT                                /FTId=VAR_002273.
SQ   SEQUENCE   763 AA;  82959 MW;  98D7881F7CB5949B CRC64;

  Query Match             75.6%;  Score 34;  DB 1;  Length 763;
  Best Local Similarity   55.6%;  Pred. No. 24;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YLSGADINL 9
          :::|||||:
Db     98 FIAGADINM 106


RESULT   6
ECHA_PIG
ID   ECHA_PIG       STANDARD;      PRT;   763 AA.
AC   Q29554;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   MITOCHONDRIAL TRIFUNCTIONAL ENZYME ALPHA SUBUNIT PRECURSOR (TP-ALPHA)
DE   (78 KDA GASTRIN-BINDING PROTEIN) [INCLUDES: LONG-CHAIN ENOYL-COA
DE   HYDRATASE (EC 4.2.1.17); LONG CHAIN 3-HYDROXYACYL-COA DEHYDROGENASE
DE   (EC 1.1.1.35)].
GN   HADHA OR LCHYD-HAD.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=STOMACH;
RX   MEDLINE; 94002212.
RA   Mantamadiotis T., Sobieszczuk P., Weinstock J., Baldwin G.S.;
RT   "Nucleotide sequence encoding a novel member of the
RT   hydratase/dehydrogenase family.";
RL   Biochim. Biophys. Acta 1170:211-215(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RX   MEDLINE; 94128080.
RA   Yang S.-Y., He X.-Y., Styles J., Luo M.J., Schulz H., Elzinga M.;
RT   "Primary structure of the large subunit of trifunctional beta-
RT   oxidation complex from pig heart mitochondria.";
RL   Biochem. Biophys. Res. Commun. 198:431-437(1994).
RN   [3]
RP   PARTIAL SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 93290643.
RA   Baldwin G.S., Casey A., Weinstock J.;
RT   "Partial structure of the gene encoding the 78 kDa gastrin binding
RT   protein excludes a close relationship with the peroxisomal
RT   trifunctional enzyme.";
RL   Biochem. Biophys. Res. Commun. 193:560-564(1993).
RN   [4]
RP   CHARACTERIZATION.
RC   TISSUE=HEART;
RX   MEDLINE; 95187623.
RA   Yang S.-Y.;
RT   "The large subunit of the pig heart mitochondrial membrane-bound beta-
RT   oxidation complex is a long-chain enoyl-CoA hydratase: 3-hydroxyacyl-
RT   CoA dehydrogenase bifunctional enzyme.";
RL   Comp. Biochem. Physiol. 109B:557-566(1994).
CC   -!- FUNCTION: BIFUNCTIONAL SUBUNIT; CANNOT USE CROTONYL-COA OR 3-
CC       HYDROXYBUTYRYL-COA AS SUBSTRATE.
CC   -!- CATALYTIC ACTIVITY: (3S)-3-HYDROXYACYL-COA = TRANS-2(OR 3)-ENOYL-
```

```
CC       COA + H(2)O.
CC   -!- CATALYTIC ACTIVITY: L-3-HYDROXYACYL-COA + NAD(+) = 3-OXOACYL-COA +
CC       NADH.
CC   -!- PATHWAY: CATALYZES TWO STEPS OF THE FATTY ACID BETA-OXIDATION
CC       CYCLE.
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE ENOYL-COA
CC       HYDRATASE/ISOMERASE FAMILY.
CC   -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE 3-
CC       HYDROXYACYL-COA DEHYDROGENASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L12581; AAA03733.1; -.
DR   EMBL; AF028609; AAB84118.1; -.
DR   HSSP; P14604; 2DUB.
DR   INTERPRO; IPR001753; -.
DR   INTERPRO; IPR002135; -.
DR   PFAM; PF00725; 3HCDH; 1.
DR   PFAM; PF00378; ECH; 1.
DR   PROSITE; PS00067; 3HCDH; 1.
DR   PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
KW   Fatty acid metabolism; Multifunctional enzyme; Oxidoreductase; NAD;
KW   Lyase; Mitochondrion; Transit peptide.
FT   TRANSIT       1     36       MITOCHONDRION (POTENTIAL).
FT   CHAIN        37    763       MITOCHONDRIAL TRIFUNCTIONAL ENZYME ALPHA
FT                                SUBUNIT.
FT   ACT_SITE    151    151       ACTIVATES A WATER MOLECULE (BY
FT                                SIMILARITY).
FT   ACT_SITE    173    173       PROVIDES THE ALPHA-PROTON (BY
FT                                SIMILARITY).
SQ   SEQUENCE   763 AA;  83106 MW;  DD0C6E7AE6B3D0F4 CRC64;


  Query Match             75.6%;  Score 34;  DB 1;  Length 763;
  Best Local Similarity   55.6%;  Pred. No. 24;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
           :::|||||:
Db      98 FIAGADINM 106


RESULT   7
WA_EMENI
ID   WA_EMENI       STANDARD;      PRT;  1986 AA.
AC   Q03149;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   CONIDIAL GREEN PIGMENT SYNTHASE (EC 2.3.1.-).
GN   WA.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae; Emericella.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93101122.
RA   Mayorga M.E., Timberlake W.E.;
RT   "The developmentally regulated Aspergillus nidulans wA gene encodes a
RT   polypeptide homologous to polyketide and fatty acid synthases.";
RL   Mol. Gen. Genet. 235:205-212(1992).
CC   -!- FUNCTION: THIS PROTEIN CONDENSES CARBON UNITS TO FORM AN
CC       INTERMEDIATE YELLOW POLYKETIDE PIGMENT THAT IS POLYMERIZED
CC       BY CONIDIAL LACCASE TO FORM THE GREEN PIGMENT IN MATURE
CC       ASEXUAL SPORES (CONIDIA).
CC   -!- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES
```

```
CC       (POTENTIAL).
CC   -!- PATHWAY: BIOSYNTHESIS OF CONIDIAL GREEN PIGMENT.
CC   -!- SIMILARITY: WITH BOTH EUKARYOTIC AND PROKARYOTIC POLYKETIDE
CC       SYNTHASES AND VERTEBRATE FATTY ACID SYNTHASES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X65866; CAA46695.1; -.
DR   PIR; S28353; S28353.
DR   INTERPRO; IPR000255; -.
DR   INTERPRO; IPR000794; -.
DR   INTERPRO; IPR001227; -.
DR   PFAM; PF00698; Acyl_transf; 1.
DR   PFAM; PF00109; ketoacyl-synt; 1.
DR   PFAM; PF00550; pp-binding; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN      529    582       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   DOMAIN      991   1024       ACYL/MALONYL TRANSFERASES
FT                                (BY SIMILARITY).
FT   DOMAIN     1650   1719       ACYL CARRIER (ACP).
FT   DOMAIN     1772   1841       ACYL CARRIER (ACP).
FT   ACT_SITE    548    548       BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT   ACT_SITE   1001   1001       ACYL/MALONYL TRANSFERASES (BY
FT                                SIMILARITY).
FT   BINDING    1682   1682       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    1804   1804       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   1986 AA;  216634 MW;  74EF0940FF40EE9A CRC64;


  Query Match             75.6%;  Score 34;  DB 1;  Length 1986;
  Best Local Similarity   75.0%;  Pred. No. 66;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGADIN 8
           ||:| |||
Db    1227 YLAGVDIN 1234


RESULT   8
V311_ASFB7
ID   V311_ASFB7     STANDARD;      PRT;   311 AA.
AC   P23163;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   D'311 PROTEIN.
GN   DP311R.
OS   African swine fever virus (strain BA71V) (ASFV).
OC   Viruses; dsDNA viruses, no RNA stage;
OC   African swine fever-like viruses.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90219205.
RA   Gonzalez A., Calvo V., Almazan F., Almendral J.M., Ramirez J.C.,
RA   de la Vega I., Blasco R., Vinuela E.;
RT   "Multigene families in African swine fever virus: family 360.";
RL   J. Virol. 64:2073-2081(1990).
RN   [2]
RP   COMPLETE GENOME.
RA   Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,
RA   Rodriguez J.F., Vinuela E.;
RT   "Analysis of the complete nucleotide sequence of African swine fever
RT   virus.";
RL   Virology 208:249-278(1995).
```

```
CC   -!- FUNCTION: NOT KNOWN.
CC   -!- SIMILARITY: BELONGS TO THE MULTIGENE FAMILY 360.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M57544; AAA42677.1; -.
DR   EMBL; U18466; AAA65374.1; -.
DR   PIR; E43680; E43680.
DR   INTERPRO; IPR002595; -.
DR   PFAM; PF01671; ASFV_360; 1.
KW   Multigene family.
SQ   SEQUENCE   311 AA;  35574 MW;  02651F4C5F2A3DA7 CRC64;

  Query Match             73.3%;  Score 33;  DB 1;  Length 311;
  Best Local Similarity   75.0%;  Pred. No. 15;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGADIN 8
             |: |||||
Db     258 YILGADIN 265


RESULT   9
ECHA_RAT
ID   ECHA_RAT       STANDARD;      PRT;   763 AA.
AC   Q64428;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   MITOCHONDRIAL TRIFUNCTIONAL ENZYME ALPHA SUBUNIT PRECURSOR (TP-ALPHA)
DE   [INCLUDES: LONG-CHAIN ENOYL-COA HYDRATASE (EC 4.2.1.17); LONG CHAIN
DE   3-HYDROXYACYL-COA DEHYDROGENASE (EC 1.1.1.35)].
GN   HADHA.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR;
RX   MEDLINE; 94075334.
RA   Kamijo T., Aoyama T., Miyazaki J., Hashimoto T.;
RT   "Molecular cloning of the cDNAs for the subunits of rat mitochondrial
RT   fatty acid beta-oxidation multienzyme complex. Structural and
RT   functional relationships to other mitochondrial and peroxisomal beta-
RT   oxidation enzymes.";
RL   J. Biol. Chem. 268:26452-26460(1993).
CC   -!- FUNCTION: BIFUNCTIONAL SUBUNIT.
CC   -!- CATALYTIC ACTIVITY: (3S)-3-HYDROXYACYL-COA = TRANS-2(OR 3)-ENOYL-
CC       COA + H(2)O.
CC   -!- CATALYTIC ACTIVITY: L-3-HYDROXYACYL-COA + NAD(+) = 3-OXOACYL-COA +
CC       NADH.
CC   -!- PATHWAY: CATALYZES TWO STEPS OF THE FATTY ACID BETA-OXIDATION
CC       CYCLE.
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE ENOYL-COA
CC       HYDRATASE/ISOMERASE FAMILY.
CC   -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE 3-
CC       HYDROXYACYL-COA DEHYDROGENASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D16478; BAA03939.1; -.
DR   HSSP; P14604; 2DUB.
DR   INTERPRO; IPR001753; -.
DR   INTERPRO; IPR002135; -.
DR   PFAM; PF00725; 3HCDH; 1.
DR   PFAM; PF00378; ECH; 1.
DR   PROSITE; PS00067; 3HCDH; 1.
DR   PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
KW   Fatty acid metabolism; Multifunctional enzyme; Oxidoreductase; NAD;
KW   Lyase; Mitochondrion; Transit peptide.
FT   TRANSIT       1     36       MITOCHONDRION (POTENTIAL).
FT   CHAIN        37    763       MITOCHONDRIAL TRIFUNCTIONAL ENZYME ALPHA
FT                                SUBUNIT.
FT   ACT_SITE    151    151       ACTIVATES A WATER MOLECULE (BY
FT                                SIMILARITY).
FT   ACT_SITE    173    173       PROVIDES THE ALPHA-PROTON (BY
FT                                SIMILARITY).
SQ   SEQUENCE   763 AA;  82512 MW;  F4A489C360AFF4A6 CRC64;

  Query Match             73.3%;  Score 33;  DB 1;  Length 763;
  Best Local Similarity   55.6%;  Pred. No. 38;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
             :::|||||:
Db      98 FVAGADINM 106


RESULT  10
YL86_YEAST
ID   YL86_YEAST     STANDARD;      PRT;   880 AA.
AC   Q06708;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL 99.8 KDA PROTEIN IN IKI3-RPS29A INTERGENIC REGION.
GN   YLR386W OR L3502.1.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE; 97313267.
RA   Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W.,
RA   Benes V., Bruckner M., Delius H., Dubois E., Dusterhoft A.,
RA   Entian K.D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA   Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Kotter P.,
RA   Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Mostl D.,
RA   Muller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA   Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA   Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA   Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,
RA   Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA   Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL   Nature 387:87-90(1997).
CC   -!- SIMILARITY: TO S.POMBE SPBC25H2.03.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U19104; AAB67272.1; -.
DR   PROSITE; PS50077; HEAT_REPEAT; 1.
KW   Hypothetical protein; Transmembrane.
```

```
FT   TRANSMEM     64     84       POTENTIAL.
SQ   SEQUENCE   880 AA;  99771 MW;  E625AB6F032D2E1B CRC64;


  Query Match             73.3%;  Score 33;  DB 1;  Length 880;
  Best Local Similarity   66.7%;  Pred. No. 45;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSGADINL 9
            |: | ||||
Db      380 YIPGQDINL 388


RESULT  11
PGL1_COLLN
ID   PGL1_COLLN     STANDARD;      PRT;   363 AA.
AC   Q00446;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   ENDOPOLYGALACTURONASE 1 PRECURSOR (EC 3.2.1.15) (PECTINASE).
GN   PG1.
OS   Colletotrichum lindemuthianum (Anthracnose fungus).
OC   Eukaryota; Fungi; Ascomycota; Phyllachorales; Phyllachoraceae;
OC   anamorphic Phyllachoraceae; Colletotrichum.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96200868.
RA   Centis S., Dumas B., Fournier J., Marolda M., Esquerre-Tugaye M.T.;
RT   "Isolation and sequence analysis of Clpg1, a gene coding for an
RT   endopolygalacturonase of the phytopathogenic fungus Colletotrichum
RT   lindemuthianum.";
RL   Gene 170:125-129(1996).
CC   -!- FUNCTION: INVOLVED IN MACERATION AND SOFT-ROTTING OF PLANT TISSUE.
CC   -!- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-ALPHA-D-
CC       GALACTOSIDURONIC LINKAGES IN PECTATE AND OTHER GALACTURONANS.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES
CC       (POLYGALACTURONASES).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X89370; CAA61552.1; -.
DR   INTERPRO; IPR000743; -.
DR   PFAM; PF00295; Glyco_hydro_28; 1.
DR   PROSITE; PS00502; POLYGALACTURONASE; 1.
KW   Hydrolase; Glycosidase; Cell wall; Signal; Zymogen; Multigene family;
KW   Glycoprotein.
FT   SIGNAL        1     17       POTENTIAL.
FT   PROPEP       18     26       POTENTIAL.
FT   CHAIN        27    363       ENDOPOLYGALACTURONASE 1.
FT   ACT_SITE    224    224       PROBABLE.
FT   CARBOHYD    212    212       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   363 AA;  36712 MW;  A17A60386791B897 CRC64;


  Query Match             71.1%;  Score 32;  DB 1;  Length 363;
  Best Local Similarity   71.4%;  Pred. No. 28;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADI 7
            |:||||:
Db      192 YISGADV 198


RESULT  12
PRS6_CAEEL
```

```
ID   PRS6_CAEEL     STANDARD;      PRT;   414 AA.
AC   P46502;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   PROBABLE 26S PROTEASE REGULATORY SUBUNIT 6B.
GN   F23F12.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Du Z.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC       DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC       COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC       26S COMPLEX (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U12965; AAA20608.1; -.
DR   WORMPEP; F23F12.6; CE01253.
DR   INTERPRO; IPR001939; -.
DR   PFAM; PF00004; AAA; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Hypothetical protein; Proteasome; ATP-binding; Nuclear protein.
FT   NP_BIND     202    209       ATP (POTENTIAL).
SQ   SEQUENCE   414 AA;  46358 MW;  8630AFFB8A2C7F32 CRC64;


  Query Match             71.1%;  Score 32;  DB 1;  Length 414;
  Best Local Similarity   85.7%;  Pred. No. 32;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSGADIN 8
            :||||||
Db      366 ISGADIN 372


RESULT  13
PRS6_HUMAN
ID   PRS6_HUMAN     STANDARD;      PRT;   418 AA.
AC   P43686;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   26S PROTEASE REGULATORY SUBUNIT 6B (TAT-BINDING PROTEIN-7) (TBP-7).
GN   PSMC4 OR TBP7.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93126329.
RA   Ohana B., Moore P.A., Ruben S.M., Southgate C.D., Green M.R.,
RA   Rosen C.A.;
RT   "The type 1 human immunodeficiency virus Tat binding protein is a
RT   transcriptional activator belonging to an additional family of
RT   evolutionarily conserved genes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:138-142(1993).
RN   [2]
RP   PARTIAL SEQUENCE, REVISION TO C-TERMINUS, AND FUNCTION.
RX   MEDLINE; 94338582.
```

```
RA   Dubiel W., Ferrell K., Rechsteiner M.;
RT   "Tat-binding protein 7 is a subunit of the 26S protease.";
RL   Biol. Chem. Hoppe-Seyler 375:237-240(1994).
CC   -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC       DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC       COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC       26S COMPLEX.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
DR   MIM; 602707; -.
DR   INTERPRO; IPR001939; -.
DR   PFAM; PF00004; AAA; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; ATP-binding; Nuclear protein.
FT   NP_BIND     206    213       ATP (POTENTIAL).
SQ   SEQUENCE   418 AA;  47336 MW;  932C5233C0AD8F08 CRC64;

  Query Match             71.1%;  Score 32;  DB 1;  Length 418;
  Best Local Similarity   85.7%;  Pred. No. 33;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSGADIN 8
            :||||||
Db      370 ISGADIN 376


RESULT 14
PRS6_MOUSE
ID   PRS6_MOUSE     STANDARD;      PRT;   418 AA.
AC   P54775;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   26S PROTEASE REGULATORY SUBUNIT 6B (CIP21) (MIP224) (MB67 INTERACTING
DE   PROTEIN).
GN   PSMC4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 96184184.
RA   Choi H.S., Seol W., Moore D.D.;
RT   "A component of the 26S proteasome binds on orphan member of the
RT   nuclear hormone receptor superfamily.";
RL   J. Steroid Biochem. Mol. Biol. 56:23-30(1996).
CC   -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC       DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC       COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC       26S COMPLEX. INTERACTS WITH THE ORPHAN NUCLEAR HORMONE RECEPTOR
CC       MB67.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L76223; AAA88243.1; -.
DR   INTERPRO; IPR001939; -.
DR   PFAM; PF00004; AAA; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; ATP-binding; Nuclear protein.
FT   NP_BIND     206    213       ATP (POTENTIAL).
SQ   SEQUENCE   418 AA;  47281 MW;  C2D57FB9591729F4 CRC64;

  Query Match             71.1%;  Score 32;  DB 1;  Length 418;
  Best Local Similarity   85.7%;  Pred. No. 33;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSGADIN 8
            :||||||
Db      370 ISGADIN 376


RESULT 15
PRS6_RAT
ID   PRS6_RAT       STANDARD;      PRT;   418 AA.
AC   Q63570;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   26S PROTEASE REGULATORY SUBUNIT 6B (TAT-BINDING PROTEIN-7) (TBP-7).
GN   PSMC4.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RX   MEDLINE; 96183075.
RA   Makino Y., Yogosawa S., Kanemaki M., Yoshida T., Yamano K.,
RA   Kishimoto T., Moncollin V., Egly J.M., Muramatsu M., Tamura T.;
RT   "Structures of the rat proteasomal ATPases: determination of highly
RT   conserved structural motifs and rules for their spacing.";
RL   Biochem. Biophys. Res. Commun. 220:1049-1054(1996).
CC   -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC       DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC       COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC       26S COMPLEX.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D50695; BAA09340.1; -.
DR   INTERPRO; IPR001939; -.
DR   PFAM; PF00004; AAA; 1.
DR   PROSITE; PS00674; AAA; 1.
KW   Proteasome; ATP-binding; Nuclear protein.
FT   NP_BIND     206    213       ATP (POTENTIAL).
SQ   SEQUENCE   418 AA;  47408 MW;  6FD2C1E3EB1FF5F0 CRC64;

  Query Match             71.1%;  Score 32;  DB 1;  Length 418;
  Best Local Similarity   85.7%;  Pred. No. 33;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSGADIN 8
            :||||||
Db      370 ISGADIN 376


Search completed: December 16, 2000, 03:05:20
Job time: 8040 sec
```

```
                    GenCore version 4.5
             Copyright (c) 1993 - 2000  Compugen Ltd.


OM protein - protein search, using sw model

Run on:          December 16, 2000, 01:51:14 ; Search time 91.14 Seconds
                                              (without alignments)
                                              6.267 Million cell updates/sec

Title:           US-09-529-121-3
Perfect score:   45
Sequence:        1 YLSGADINL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters:      182106

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_65:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                 %
Result          Query
  No.   Score   Match Length DB  ID                   Description
-------------------------------------------------------------------------------
    1      38    84.4    702  2  A36319               carcinoembryonic a
    2      34    75.6    176  2  F71064               micrococcal nuclea
    3      34    75.6    275  2  PN0511               gastrin-binding pr
    4      34    75.6    714  2  C65007               hypothetical prote
    5      34    75.6    763  2  JC2108               long-chain-fatty-a
    6      34    75.6    860  2  C72338               hypothetical prote
    7      34    75.6   1986  2  S28353               probable polyketid
    8      33    73.3    311  2  E43680               D'311 protein - Af
    9      33    73.3    401  2  T24929               hypothetical prote
   10      33    73.3    402  2  T24930               hypothetical prote
   11      33    73.3    441  2  T20840               hypothetical prote
   12      33    73.3    445  2  T21262               hypothetical prote
   13      33    73.3    454  2  T21261               hypothetical prote
   14      33    73.3    705  2  T12784               sublancin 168 lant
   15      33    73.3    763  1  A49681               long-chain-fatty-a
   16      33    73.3    880  2  S51473               probable membrane
   17      32    71.1    188  2  G70475               conserved hypothet
   18      32    71.1    312  2  T33344               hypothetical prote
   19      32    71.1    350  2  T22450               hypothetical prote
   20      32    71.1    363  2  JC4748               polygalacturonase
   21      32    71.1    374  2  E69049               hypothetical prote
   22      32    71.1    506  2  C81704               monooxygenase-rela
   23      32    71.1    540  2  E69861               ABC transporter (A
   24      32    71.1   1023  2  T30257               IgG Fc binding pro
   25      32    71.1   1650  2  S53457               dominant autoantig
   26      32    71.1   4660  2  T42737               gp330 protein prec
   27      31    68.9     30  2  A03148               retinol-binding pr
   28      31    68.9    108  2  S37313               hlyU protein - Vib
   29      31    68.9    112  2  PQ0493               hypothetical prote
```

```
   30      31    68.9    172  1  A41841               phycoerythrocyanin
   31      31    68.9    175  2  D75083               micrococcal nuclea
   32      31    68.9    210  2  S08389               hypothetical prote
   33      31    68.9    301  2  H75470               histone deacetylas
   34      31    68.9    491  2  A49179               melanoma antigen h
   35      31    68.9    502  2  T02746               cyclin A-like prot
   36      31    68.9    502  2  S50519               hypothetical prote
   37      31    68.9    554  2  A70904               probable acid--CoA
   38      31    68.9    586  2  S19381               hypothetical prote
   39      31    68.9    596  2  T23685               hypothetical prote
   40      31    68.9    597  2  A30814               development-specif
   41      31    68.9    699  2  B71723               translation elonga
   42      31    68.9    854  1  QRHYLD               LDL receptor precu
   43      31    68.9    856  2  T10585               serine proteinase
   44      31    68.9    860  1  QRHULD               LDL receptor precu
   45      31    68.9    862  1  QRMSLD               LDL receptor precu


                              ALIGNMENTS


RESULT   1
A36319
carcinoembryonic antigen precursor - human
N;Alternate names: CEA; meconium antigen 100
C;Species: Homo sapiens (man)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 31-Jan-2000
C;Accession: A36319; A27773; A31037; A25845; S08106; S31737; A44476; I54224; I59098;
R;Schrewe, H.; Thompson, J.; Bona, M.; Hefta, L.J.F.; Maruya, A.; Hassauer, M.; Shive
Mol. Cell. Biol. 10, 2738-2748, 1990
A;Title: Cloning of the complete gene for carcinoembryonic antigen: analysis of its p
A;Reference number: A36319; MUID:90258861
A;Accession: A36319
A;Molecule type: DNA
A;Residues: 1-702 <SCH>
A;Cross-references: GB:M17303; NID:g178676; PIDN:AAB59513.1; PID:g178677
A;Note: the authors show the codons TTA for residue 641-Phe and CAG for residue 646-T
R;Beauchemin, N.; Benchimol, S.; Cournoyer, D.; Fuks, A.; Stanners, C.P.
Mol. Cell. Biol. 7, 3221-3230, 1987
A;Title: Isolation and characterization of full-length functional cDNA clones for hum
A;Reference number: A27773; MUID:88038876
A;Accession: A27773
A;Molecule type: mRNA
A;Residues: 1-702 <BEA>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
R;Barnett, T.; Goebel, S.J.; Nothdurft, M.A.; Elting, J.J.
Genomics 3, 59-66, 1988
A;Title: Carcinoembryonic antigen family: characterization of cDNAs coding for NCA an
A;Reference number: A31037; MUID:89122014
A;Accession: A31037
A;Molecule type: mRNA
A;Residues: 1-702 <BAR>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
A;Note: the authors translated the codon GTG for residue 130 as Leu
R;Oikawa, S.; Nakazato, H.; Kosaki, G.
Biochem. Biophys. Res. Commun. 142, 511-518, 1987
A;Title: Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA
A;Reference number: A25845; MUID:87128144
A;Accession: A25845
A;Molecule type: mRNA
A;Residues: 5-702 <OIK>
A;Cross-references: GB:M15042; NID:g180198; PIDN:AAA51963.1; PID:g180199
R;Oikawa, S.
submitted to the EMBL Data Library, September 1989
A;Reference number: S08106
A;Accession: S08106
A;Molecule type: mRNA
A;Residues: 5-319,321-702 <OI2>
A;Cross-references: EMBL:X16455; NID:g29854; PIDN:CAA34474.1; PID:g825638
R;Barnett, T.
submitted to the EMBL Data Library, September 1991
A;Description: Genomic DNA sequence upstream of the translational start of the carcin
A;Reference number: S31737
```

```
A;Accession: S31737
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-141 <BA2>
A;Cross-references: EMBL:X62151
R;Khan, W.N.; Fraengsmyr, L.; Teglund, S.; Israelsson, A.; Bremer, K.; Hammarstroem, S.
Genomics 14, 384-390, 1992
A;Title: Identification of three new genes and estimation of the size of the carcinoembr
A;Reference number: A44476; MUID:93052339
A;Accession: A44476
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 35-141 <KHA>
R;Willcocks, T.C.; Craig, I.W.
Genomics 8, 492-500, 1990
A;Title: Characterization of the genomic organization of human carcinoembryonic antigen
A;Reference number: I54224; MUID:91139118
A;Accession: I54224
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-37 <RES>
A;Cross-references: GB:M60964; NID:g180215; PIDN:AAA51964.1; PID:g180217
R;Zimmermann, W.; Ortlieb, B.; Friedrich, R.; von Kleist, S.
Proc. Natl. Acad. Sci. U.S.A. 84, 2960-2964, 1987
A;Title: Isolation and characterization of cDNA clones encoding the human carcinoembryon
A;Reference number: I59098; MUID:87204247
A;Accession: I59098
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 331-702 <RE2>
A;Cross-references: GB:M16234; NID:g180240; PIDN:AAA51972.1; PID:g180241
R;Siepen, D.; Paxton, R.J.; Neumaier, M.; Shively, J.E.; Wagener, C.
Biochem. Biophys. Res. Commun. 147, 212-218, 1987
A;Title: Carcinoembryonic antigen (CEA) and two crossreacting antigens of 165 KD and 105
A;Reference number: A26831; MUID:87326349
A;Accession: A26831
A;Molecule type: protein
A;Residues: 35-64 <SIE>
R;Thomas, P.; Toth, C.A.
Biochem. Biophys. Res. Commun. 170, 391-396, 1990
A;Title: Carcinoembryonic antigen binding to Kupffer cells is via a peptide located at t
A;Reference number: A35490; MUID:90321257
A;Accession: A35490
A;Molecule type: protein
A;Residues: 'X',140-151,'X',153,'X',155-156 <THO>
A;Note: this is the amino terminal end of a fragment shown to mediate uptake by Kupffer
C;Comment: This heavily glycosylated membane protein of unknown function is a widely use
C;Comment: This protein may be processed at its C-terminus. It is anchored to the membra
C;Genetics:
A;Gene: GDB:CEA
A;Cross-references: GDB:119054; OMIM:114890
A;Map position: 19q13.2-19q13.2
A;Introns: 22/1; 142/1; 235/1; 320/1; 413/1; 498/1; 591/1; 676/1
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin
C;Keywords: blocked carboxyl end; glycoprotein; lipoprotein; membrane protein; phosphati
F;1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-678/Product: carcinoembryonic antigen #status predicted <MAT>
F;160-217/Domain: immunoglobulin homology <IMM1>
F;252-301/Domain: immunoglobulin homology <IMM2>
F;338-395/Domain: immunoglobulin homology <IMM3>
F;516-573/Domain: immunoglobulin homology <IMM4>
F;608-657/Domain: immunoglobulin homology <IMM5>
F;679-702/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F;678/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Gly) (in mature form


  Query Match             84.4%;  Score 38;  DB 2;  Length 702;
  Best Local Similarity   77.8%;  Pred. No. 8.6;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADINL 9
            |||||::||
Db      605 YLSGANLNL 613


RESULT   2
F71064
micrococcal nuclease (EC 3.1.31.1) PH1212 precursor - Pyrococcus horikoshii
N;Alternate names: thermonuclease homolog
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C;Accession: F71064
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophili
A;Reference number: A71000; MUID:98344137
A;Accession: F71064
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-176 <KAW>
A;Cross-references: GB:AP000005; NID:g3236132; PIDN:BAA30312.1; PID:g3257629
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBa
C;Genetics:
A;Gene: PH1212
C;Superfamily: micrococcal nuclease
C;Keywords: hydrolase
F;1-27/Domain: signal sequence #status predicted <SIG>


  Query Match             75.6%;  Score 34;  DB 2;  Length 176;
  Best Local Similarity   75.0%;  Pred. No. 13;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGADIN 8
            ||:| |||
Db      128 YLNGTDIN 135


RESULT   3
PN0511
gastrin-binding protein precursor - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 17-Mar-1999
C;Accession: PN0511
R;Baldwin, G.S.; Casey, A.; Weinstock, J.
Biochem. Biophys. Res. Commun. 193, 560-564, 1993
A;Title: Partial structure of the gene encoding the 78KDa gastrin binding protein enc
A;Reference number: PN0511; MUID:93290643
A;Accession: PN0511
A;Molecule type: mRNA
A;Residues: 1-275 <BAL>
A;Note: complete nucleotide sequence not given
C;Genetics:
A;Introns: 23/1; 37/1; 60/3; 105/2; 151/3; 191/3; 226/1
C;Superfamily: enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; 3-hydroxyacyl-CoA
F;1-36/Domain: signal sequence #status predicted <SIG>
F;37-275/Product: gastrin-binding protein #status predicted <MAT>
F;62-218/Domain: enoyl-CoA hydratase homology <ECH>


  Query Match             75.6%;  Score 34;  DB 2;  Length 275;
  Best Local Similarity   55.6%;  Pred. No. 21;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADINL 9
            :::|||||:
Db       98 FIAGADINM 106


RESULT   4
C65007
hypothetical protein b2341 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
```

```
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 22-Jun-1999
C;Accession: C65007
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: C65007
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-714 <BLAT>
A;Cross-references: GB:AE000322; GB:U00096; NID:g1788672; PIDN:AAC75401.1; PID:g1788682;
A;Experimental source: strain K-12, substrain MG1655
C;Superfamily: enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; 3-hydroxyacyl-CoA de
F;29-185/Domain: enoyl-CoA hydratase homology <ECH>
F;308-587/Domain: 3-hydroxyacyl-CoA dehydrogenase homology <HCD>


  Query Match             75.6%;  Score 34;  DB 2;  Length 714;
  Best Local Similarity   55.6%;  Pred. No. 58;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              :::|||||:
Db         65 FIAGADINM 73


RESULT    5
JC2108
long-chain-fatty-acid beta-oxidation multienzyme complex alpha chain precursor, mitochon
N;Alternate names: 78K gastrin-binding protein
N;Contains: long-chain-3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.211); long-chain-enoyl-
C;Species: Homo sapiens (man)
C;Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 20-Jun-2000
C;Accession: JC2108; PC2058; S50127
R;Kamijo, T.; Aoyama, T.; Komiyama, A.; Hashimoto, T.
Biochem. Biophys. Res. Commun. 199, 818-825, 1994
A;Title: Structural analysis of cDNAs for subunits of human mitochondrial fatty acid bet
A;Reference number: JC2108; MUID:94183263
A;Accession: JC2108
A;Molecule type: mRNA
A;Residues: 1-763 <KAM>
A;Cross-references: GB:D16480; NID:g493657; PIDN:BAA03941.1; PID:g862457
A;Accession: PC2058
A;Molecule type: protein
A;Residues: 37-64 <KA2>
A;Note: peptide sequence from amino end of mature protein
R;Zhang, Q.X.; Baldwin, G.S.
Biochim. Biophys. Acta 1219, 567-575, 1994
A;Title: Structures of the human cDNA and gene encoding the 78 kDa gastrin-binding prote
A;Reference number: S50127; MUID:95002180
A;Accession: S50127
A;Status: preliminary; translation not shown
A;Molecule type: mRNA
A;Residues: 1-145,'L',147-151,'L',153-170,'A',172-177,'I',179-196,'VF',199-205,'N',207-2
A;Cross-references: EMBL:U04627; NID:g595266; PIDN:AAA56664.1; PID:g595267
C;Genetics:
A;Gene: GDB:HADHA
A;Cross-references: GDB:434026; OMIM:600890
A;Map position: 2p23-2p23
C;Complex: heterooctamer of 4 alpha and 4 beta chains
C;Superfamily: enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; 3-hydroxyacyl-CoA de
C;Keywords: acyltransferase; carbon-oxygen lyase; fatty acid beta-oxidation; heterooctam
F;1-36/Domain: transit peptide (mitochondrion) #status predicted <TPP>
F;37-763/Product: fatty acid beta-oxidation trifunctional protein, alpha chain #status e
F;62-218/Domain: enoyl-CoA hydratase homology <ECH>
F;361-640/Domain: 3-hydroxyacyl-CoA dehydrogenase homology <HCD>
F;363-391/Region: beta-alpha-beta NAD nucleotide-binding fold


  Query Match             75.6%;  Score 34;  DB 2;  Length 763;
  Best Local Similarity   55.6%;  Pred. No. 62;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 YLSGADINL 9
              :::|||||:
Db         98 FIAGADINM 106


RESULT    6
C72338
hypothetical protein - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: C72338
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
 C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A;Reference number: A72200; MUID:99287316
A;Accession: C72338
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-860 <ARN>
A;Cross-references: GB:AE001745; GB:AE000512; NID:g4981278; PID:g4981284; TIGR:TM0757
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0757


  Query Match             75.6%;  Score 34;  DB 2;  Length 860;
  Best Local Similarity   75.0%;  Pred. No. 70;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADIN 8
              ||:| |||
Db        443 YLTGGDIN 450


RESULT    7
S28353
probable polyketide synthase - Emericella nidulans
C;Species: Emericella nidulans, Aspergillus nidulans
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 26-May-2000
C;Accession: S28353
R;Mayorga, M.E.; Timberlake, W.E.
Mol. Gen. Genet. 235, 205-212, 1992
A;Title: The developmentally regulated Aspergillus nidulans wA gene encodes a polypep
A;Reference number: S28353; MUID:93101122
A;Accession: S28353
A;Molecule type: DNA
A;Residues: 1-1986 <MAY>
A;Cross-references: EMBL:X65866; NID:g5508; PID:g5509
C;Genetics:
A;Gene: wA
A;Introns: 96/2; 193/3; 1336/3; 1588/3
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier pro
C;Keywords: carrier protein
F;397-805/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;911-1199/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1648-1718/Domain: acyl carrier protein homology <ACP>
F;1766-1840/Domain: acyl carrier protein homology <ACP1>


  Query Match             75.6%;  Score 34;  DB 2;  Length 1986;
  Best Local Similarity   75.0%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADIN 8
              ||:| |||
Db       1227 YLAGVDIN 1234


RESULT    8
E43680
```

```
D'311 protein - African swine fever virus (strain BA71V)
C;Species: African swine fever virus, ASFV
C;Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 26-Aug-1999
C;Accession: E43680
R;Gonzalez, A.; Calvo, V.; Almazan, F.; Almendral, J.M.; Ramirez, J.C.; De La Vega, I.;
J. Virol. 64, 2073-2081, 1990
A;Title: Multigene families in African swine fever virus: family 360.
A;Reference number: A43680; MUID:90219205
A;Accession: E43680
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-311 <GON>
A;Cross-references: GB:M57546
C;Superfamily: African swine fever virus L356 protein


  Query Match             73.3%;  Score 33;  DB 2;  Length 311;
  Best Local Similarity   75.0%;  Pred. No. 38;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADIN 8
              |: |||||
Db        258 YILGADIN 265


RESULT    9
T24929
hypothetical protein T15D6.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24929
R;Dobson, R.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19956
A;Accession: T24929
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-401 <WIL>
A;Cross-references: EMBL:Z83125; PIDN:CAB05620.1; GSPDB:GN00019; CESP:T15D6.2
A;Experimental source: clone T15D6
C;Genetics:
A;Gene: CESP:T15D6.2
A;Map position: 1
A;Introns: 45/3; 75/2; 111/3; 151/3; 227/3; 252/3; 296/2; 359/3


  Query Match             73.3%;  Score 33;  DB 2;  Length 401;
  Best Local Similarity   77.8%;  Pred. No. 50;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              |||| || |
Db        186 YLSGVDIPL 194


RESULT   10
T24930
hypothetical protein T15D6.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24930
R;Dobson, R.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19956
A;Accession: T24930
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-402 <WIL>
A;Cross-references: EMBL:Z83125; PIDN:CAB05621.1; GSPDB:GN00019; CESP:T15D6.3
A;Experimental source: clone T15D6
C;Genetics:
A;Gene: CESP:T15D6.3
A;Map position: 1
A;Introns: 55/3; 85/2; 121/3; 161/3; 237/3; 262/3

  Query Match             73.3%;  Score 33;  DB 2;  Length 402;
  Best Local Similarity   77.8%;  Pred. No. 50;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              |||| || |
Db        196 YLSGVDIPL 204


RESULT   11
T20840
hypothetical protein F13E9.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20840
R;McMurray, A.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z19332
A;Accession: T20840
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-441 <WIL>
A;Cross-references: EMBL:Z69383; PIDN:CAA93405.1; GSPDB:GN00022; CESP:F13E9.1
A;Experimental source: clone F13E9
C;Genetics:
A;Gene: CESP:F13E9.1
A;Map position: 4
A;Introns: 39/3; 93/1; 146/3; 252/3; 296/2; 426/2


  Query Match             73.3%;  Score 33;  DB 2;  Length 441;
  Best Local Similarity   62.5%;  Pred. No. 55;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADIN 8
              |:|| |:|
Db        427 YISGEDVN 434


RESULT   12
T21262
hypothetical protein F22D6.11 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T21262
R;Wilkinson, J.
submitted to the EMBL Data Library, April 1996
A;Reference number: Z19397
A;Accession: T21262
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-445 <WIL>
A;Cross-references: EMBL:Z71262; PIDN:CAA95817.1; GSPDB:GN00019; CESP:F22D6.11
A;Experimental source: clone F22D6
C;Genetics:
A;Gene: CESP:F22D6.11
A;Map position: 1
A;Introns: 65/3; 95/2; 136/3; 180/3; 256/3; 281/3; 403/3


  Query Match             73.3%;  Score 33;  DB 2;  Length 445;
  Best Local Similarity   77.8%;  Pred. No. 56;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              |||| || |
Db        215 YLSGVDIPL 223


RESULT   13
```

```
T21261
hypothetical protein F22D6.12 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T21261
R;Wilkinson, J.
submitted to the EMBL Data Library, April 1996
A;Reference number: Z19397
A;Accession: T21261
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-454 <WIL>
A;Cross-references: EMBL:Z71262; PIDN:CAA95816.1; GSPDB:GN00019; CESP:F22D6.12
A;Experimental source: clone F22D6
C;Genetics:
A;Gene: CESP:F22D6.12
A;Map position: 1
A;Introns: 62/3; 92/2; 190/3; 266/3; 291/3; 412/3


  Query Match             73.3%;  Score 33;  DB 2;  Length 454;
  Best Local Similarity   77.8%;  Pred. No. 57;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
           |||| || |
Db     225 YLSGVDIPL 233


RESULT  14
T12784
sublancin 168 lantibiotic transporter sunT - Bacillus subtilis phage SPBc2
N;Alternate names: ABC transport protein yolH
C;Species: Bacillus subtilis phage SPBc2
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 20-Jun-2000
C;Accession: T12784; A69720
R;Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Mauel, C.; Karamata, D.
submitted to the EMBL Data Library, August 1997
A;Description: The complete nucleotide sequence of the Bacillus subtilis SPbetac2 propha
A;Reference number: Z17583
A;Accession: T12784
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-705 <LAZ>
A;Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025498; PIDN:AAC12993.1
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033
A;Accession: A69720
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-705 <KUN>
A;Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14065.1; PID:g2634567
A;Experimental source: strain 168
C;Genetics: <LA>
A;Gene: yolH
C;Genetics: <KU>
A;Gene: sunT
C;Superfamily: hemolysin secretion protein B; ATP-binding cassette homology
C;Keywords: ATP; P-loop
```

```
F;499-690/Domain: ATP-binding cassette homology <ABC>
F;516-523/Region: nucleotide-binding motif A (P-loop)


  Query Match             73.3%;  Score 33;  DB 2;  Length 705;
  Best Local Similarity   75.0%;  Pred. No. 91;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGADIN 8
           ||:| |||
Db     541 YLNGLDIN 548


RESULT  15
A49681
long-chain-fatty-acid beta-oxidation multienzyme complex alpha chain precursor, mitoc
N;Contains: long-chain-3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.211); long-chain-eno
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: A49681
R;Kamijo, T.; Aoyama, T.; Miyazaki, J.; Hashimoto, T.
J. Biol. Chem. 268, 26452-26460, 1993
A;Title: Molecular cloning of the cDNAs for the subunits of rat mitochondrial fatty a
omal beta-oxidation enzymes.
A;Reference number: A49681; MUID:94075334
A;Accession: A49681
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-763 <KAM>
A;Cross-references: GB:D16478; NID:g510107; PIDN:BAA03939.1; PID:g510108
A;Experimental source: Wistar, liver
A;Note: sequence extracted from NCBI backbone (NCBIN:140844, NCBIP:140846)
C;Complex: heterooctamer of 4 alpha and 4 beta chains
C;Superfamily: enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; 3-hydroxyacyl-CoA
C;Keywords: carbon-oxygen lyase; fatty acid beta-oxidation; heterooctamer; hydro-lyas
F;62-218/Domain: enoyl-CoA hydratase homology <ECH>
F;361-640/Domain: 3-hydroxyacyl-CoA dehydrogenase homology <HCD>
F;363-391/Region: beta-alpha-beta NAD nucleotide-binding fold


  Query Match             73.3%;  Score 33;  DB 1;  Length 763;
  Best Local Similarity   55.6%;  Pred. No. 99;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
           :::|||||:
Db      98 FVAGADINM 106


Search completed: December 16, 2000, 01:51:15
Job time: 7636 sec
```

```
                  GenCore version 4.5
          Copyright (c) 1993 - 2000  Compugen Ltd.


OM protein - protein search, using sw model

Run on:          December 16, 2000, 00:51:14 ; Search time 108.84 Seconds
                                                (without alignments)
                                                2.827 Million cell updates/sec

Title:           US-09-529-121-3
Perfect score:   45
Sequence:        1 YLSGADINL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters:      268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_36:*
                1:  /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
                2:  /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
                3:  /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
                4:  /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
                5:  /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
                6:  /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
                7:  /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
                8:  /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
                9:  /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
                10:  /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
                11:  /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
                12:  /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
                13:  /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
                14:  /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
                15:  /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
                16:  /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
                17:  /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
                18:  /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
                19:  /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
                20:  /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
                21:  /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                   %
Result          Query
  No.   Score   Match Length DB  ID                       Description
----------------------------------------------------------------------------
   1      45    100.0      9  20  Y09527                    Carcinoembryonic a
   2      43     95.6      9  20  Y09526                    Carcinoembryonic a
   3      40     88.9      9  20  Y09528                    Carcinoembryonic a
   4      38     84.4      9  18  W39723                    Human carcina-embr
   5      38     84.4      9  19  W77134                    CEA synthetic pept
   6      38     84.4      9  19  W70045                    CEA derived HLA-A2
   7      38     84.4      9  20  Y47655                    Immunogenic peptid
   8      38     84.4      9  20  Y09525                    Carcinoembryonic a
   9      38     84.4     10  20  Y46555                    Immunogenic peptid
  10      38     84.4    107  20  W86133                    Protein sequence o
  11      38     84.4    178  10  P93499                    Sequence of carcin
  12      38     84.4    468  16  R77436                    BGP (1-314)/CEA (4
  13      38     84.4    493  16  R77435                    BGP (1-314)/CEA (4
  14      38     84.4    509  16  R77437                    BGP (1-314)/CEA (4
  15      38     84.4    511  16  R77438                    BGP (1-314)/CEA (4
  16      38     84.4    642  15  R60619                    Carcinoembryonic a
  17      38     84.4    663  17  R98519                    Immunogenic carcin
  18      38     84.4    698  9  P81229                    Carcinoembryonic a
  19      38     84.4    698  16  R65168                    Carcinoembryonic a
  20      38     84.4    698  18  W22844                    Human carcinoembry
  21      38     84.4    702  9  P81222                    Carcinoembryonic a
  22      38     84.4    702  10  P94014                    Carcinoembryonic c
  23      38     84.4    702  10  P93999                    Amino acid sequenc
  24      38     84.4    702  15  R54713                    Carcinoembryonic a
  25      38     84.4    702  17  W06872                    Carcinoembryonic a
  26      38     84.4    702  20  W83137                    CEA protein.  Homo
  27      38     84.4    734  17  W00182                    Carcinoembryonic a
  28      35     77.8      9  21  Y54173                    HLA binding peptid
  29      34     75.6      9  20  Y09529                    Carcinoembryonic a
  30      34     75.6     81  15  R46083                    Enoyl CoA hydrase
  31      34     75.6    572  20  Y28885                    Human glypican 5 p
  32      34     75.6    572  20  Y25918                    Human GPC5 protein
  33      34     75.6    572  20  Y25913                    Human GPC5 protein
  34      34     75.6    709  20  Y31645                    Human transport-as
  35      34     75.6    763  13  R28754                    Gastrin-binding pr
  36      33     73.3   4655  19  W43313                    Human kidney calci
  37      32     71.1    202  20  Y08444                    Soybean Pad1 homol
  38      32     71.1    307  20  Y08442                    Z. mays Pad1 homol
  39      32     71.1    309  20  Y08443                    Soybean Pad1 homol
  40      32     71.1    635  18  W25755                    PHA depolymerase.
  41      31     68.9      9  17  W00680                    Peptide comprising
  42      31     68.9    207  17  R97213                    Human calcium sens
  43      31     68.9    207  19  W26770                    Calcium sensor pro
  44      31     68.9    222  20  Y05539                    Wheat Type I gluta
  45      31     68.9    280  16  R64141                    Low density lipopr


                                   ALIGNMENTS


RESULT   1
Y09527
ID   Y09527 standard; peptide; 9 AA.
XX
AC   Y09527;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:3.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
```

```
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;


  Query Match             100.0%;  Score 45;  DB 20;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              |||||||||
Db          1 ylsgadinl 9


RESULT   2
Y09526
ID   Y09526 standard; peptide; 9 AA.
XX
AC   Y09526;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:2.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
```

```
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             95.6%;  Score 43;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              ||||||:||
Db          1 ylsgadlnl 9


RESULT   3
Y09528
ID   Y09528 standard; peptide; 9 AA.
XX
AC   Y09528;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:4.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;


  Query Match             88.9%;  Score 40;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 YLSGADINL 9
           |||||:|||
Db       1 ylsganinl 9


RESULT   4
W39723
ID   W39723 standard; peptide; 9 AA.
XX
AC   W39723;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Human carcina-embryonic antigen (CEA) peptide (pos. 571-579).
XX
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
XX
OS   Homo sapiens.
XX
PN   WO9741440-A1.
XX
PD   06-NOV-1997.
XX
PF   28-APR-1997;   97WO-NL00229.
XX
PR   23-DEC-1996;   96EP-0203670.
PR   26-APR-1996;   96EP-0201145.
XX
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
PI   Kast WM,  Melief CJM,  Offringa R,  Toes REM,  Van Der Burg SH;
XX
DR   WPI; 1997-549891/50.
XX
PT   Method of selecting T cell peptide epitope(s) - by measuring the
PT   stability of HLA class I-peptide complexes on intact B cells
XX
PS   Example 3; Page 85; 109pp; English.
XX
CC   Peptides W39430-W39734 are used in a novel method for the selection of
CC   immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC   method involves the identification of peptide sequences capable of
CC   binding to an HLA (human leukocyte antigen) class I molecule and
CC   measuring the binding of this epitope peptide to the HLA class I peptide.
CC   The stability of binding of the peptide and MHC (major histocompatablilty
CC   complex) class I molecule is measured on intact human B cells carrying
CC   the MHC molecule at their cell surfaces. The method can be used to select
CC   peptide epitopes for generating vaccines against a disease associated
CC   with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC   especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC   immune responses. Peptide W39723 is derived from the human
CC   carcino-embryonic antigen (CEA) and has the ability to bind to the human
CC   MHC Class I allele HLA-A2.1.
XX
SQ   Sequence   9 AA;


  Query Match             84.4%;  Score 38;  DB 18;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
           |||||::||
Db       1 ylsganlnl 9


RESULT   5
W77134
ID   W77134 standard; peptide; 9 AA.
XX
AC   W77134;
XX
DT   16-NOV-1998  (first entry)
XX
DE   CEA synthetic peptide epitope 1.
XX
KW   Tyrosinase; tyrosinase cytotoxic lymphocyte response;
KW   cytotoxic T lymphocyte; cysteine-depleted; melanoma.
XX
OS   Synthetic.
XX
PN   WO9833810-A2.
XX
PD   06-AUG-1998.
XX
PF   29-JAN-1998;   98WO-US01592.
XX
PR   30-JAN-1997;   97US-0037781.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI   Engelhard VH,  Hunt DF,  Kittlesen D,  Slingluff CL;
XX
DR   WPI; 1998-437388/37.
XX
PT   Disease specific immunogen  - comprises disease specific cytotoxic T
PT   lymphocyte epitope used to elicit melanoma specific CTL response
XX
PS   Disclosure; Page 27; 93pp; English.
XX
CC   The peptide epitope W77119-W77138 were created for human tumour-specific
CC   cytotoxic T lymphocyte response.  These peptides are are cysteine-
CC   depleted mutants of a native disease-specific CTL epitope.  The cysteine-
CC   depleted CTL epitopes elicit a stronger or more specific CTL response
CC   than the native epitope.  The epitopes can be used in a disease-specific
CC   immunogen to protect a mammal against disease in particular melanomas.
CC   The peptides may also be used to screen a sample for the presence of
CC   an antigen with the same epitope, or with a different cross-reactive
CC   epitope.
XX
SQ   Sequence   9 AA;


  Query Match             84.4%;  Score 38;  DB 19;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
           |||||::||
Db       1 ylsganlnl 9


RESULT   6
W70045
ID   W70045 standard; peptide; 9 AA.
XX
AC   W70045;
XX
DT   22-OCT-1998  (first entry)
XX
DE   CEA derived HLA-A2.1 binding peptide 2 (residues 605-613).
XX
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW   antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW   viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW   fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9833888-A1.
XX
```

```
PD   06-AUG-1998.
XX
PF   30-JAN-1998;   98WO-US01959.
XX
PR   31-JAN-1997;   97US-0036696.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Celis E,  Sette A,  Sidney J,  Southwood S,  Tsai V;
XX
DR   WPI; 1998-437445/37.
XX
PT   Production of antigen-specific cytotoxic T cells - by incubating
PT   immunogenic peptide(s) from antigen that binds class I major
PT   histocompatibility complex molecules with pre-treated antigen
PT   presenting cells
XX
PS   Example 6; Page 75; 104pp; English.
XX
CC   Sequences shown in W70044 to W70052 represent peptides derived from
CC   carcinoembryonic antigen (CEA). The peptides can bind to a human
CC   leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method
CC   of invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC   vitro. The method comprises contacting immunogenic peptides from an
CC   antigen that binds class I major histocompatibility complex (MHC)
CC   molecules with antigen presenting cells (APCs) pretreated with
CC   pretreatment growth factors, and incubating the APCs with purified CD8
CC   cells in the presence of at least 2 incubation growth factors, thereby
CC   producing antigen-specific CTLs. A method for specifically killing
CC   target cells in a human patient is also provided which comprises
CC   obtaining a fluid sample containing CTLs from a patient, contacting the
CC   cytotoxic T cells with APCs pretreated with pre-treatment growth factors,
CC   where the APCs comprise class I MHC molecules. The pretreated APCs are
CC   incubated with the cytotoxic growth factors, thereby producing activated
CC   CTLs which are contacted with a carrier to form a composition. The
CC   composition can then be administered to the patient. The activated CTLs
CC   can be used for treating cancers, immune disorders, viral infections,
CC   AIDS, hepatitis, bacterial infection, fungal infection, malaria or
CC   tuberculosis.
XX
SQ   Sequence   9 AA;

  Query Match             84.4%;  Score 38;  DB 19;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 YLSGADINL 9
             |||||::||
Db         1 ylsganlnl 9

RESULT   7
Y47655
ID   Y47655 standard; Peptide; 9 AA.
XX
AC   Y47655;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2266.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
```

```
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 118; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;

  Query Match             84.4%;  Score 38;  DB 20;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 YLSGADINL 9
             |||||::||
Db         1 ylsganlnl 9

RESULT   8
Y09525
ID   Y09525 standard; peptide; 9 AA.
XX
AC   Y09525;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist CAP-1.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
```

```
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 1; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;


  Query Match             84.4%;  Score 38;  DB 20;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADINL 9
            |||||::||
Db        1 ylsganlnl 9


RESULT   9
Y46555
ID   Y46555 standard; Peptide; 10 AA.
XX
AC   Y46555;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #1166.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
```

```
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 76; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   10 AA;

  Query Match             84.4%;  Score 38;  DB 20;  Length 10;
  Best Local Similarity   77.8%;  Pred. No. 0.041;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADINL 9
              |||||::||
Db        2 ylsganlnl 10


RESULT  10
W86133
ID   W86133 standard; Protein; 107 AA.
XX
AC   W86133;
XX
DT   03-MAR-1999  (first entry)
XX
DE   Protein sequence of vaccine 2 708 V1.
XX
KW   Non-immunogenic; epitope; T-cell; immunogenicity; immune system; SK;
KW   immunoglobulin; therapeutic; streptokinase; vaccine; 708.
XX
OS   Homo sapiens.
XX
PN   WO9852976-A1.
XX
PD   26-NOV-1998.
XX
PF   21-MAY-1998;   98WO-GB01473.
XX
PR   14-APR-1998;   98GB-0007751.
PR   21-MAY-1997;   97GB-0010480.
PR   31-JUL-1997;   97GB-0016197.
PR   28-NOV-1997;   97GB-0025270.
PR   02-DEC-1997;   97US-0067235.
XX
PA   (BIOV-) BIOVATION LTD.
XX
PI   Carr FJ;
XX
DR   WPI; 1999-045301/04.
XX
PT   Reducing immunogenicity of proteins - by modifying the amino acid
```

```
PT   sequence of the protein to eliminate potential epitopes for T-cells
PT   of a given species
XX
PS   Example 4; Fig 19; 77pp; English.
XX
CC   The invention relates to a method for the production of non-immunogenic
CC   proteins. The method comprises determining at least part of the amino
CC   acid sequence of the protein; (b) identifying in the amino acid sequence
CC   one or more potential epitopes for T-cells (T-cell epitopes) of the given
CC   species; and (c) modifying the amino acid sequence to eliminate at least
CC   one of the T-cell epitopes identified in step (b) thereby to eliminate or
CC   reduce the immunogenicity of the protein when exposed to the immune
CC   system of the given species. A method of analysing a pre-existing protein
CC   to predict the basis for immunogenic responses is also provided. The
CC   methods can be used particularly for reducing the immunogenicity of
CC   immunoglobulins or therapeutic proteins, e.g. Streptokinase (SK). The
CC   products can be used for diagnosis and therapy. The present sequence
CC   represents the protein sequence of vaccine 2 708 V1.
XX
SQ   Sequence   107 AA;

  Query Match             84.4%;  Score 38;  DB 20;  Length 107;
  Best Local Similarity   77.8%;  Pred. No. 0.67;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
             |||||::||
Db      89 ylsganlnl 97


RESULT 11
P93499
ID   P93499 standard; protein; 178 AA.
XX
AC   P93499;
XX
DT   08-MAY-1990  (first entry)
XX
DE   Sequence of carcinoembryonic antigen domain III.
XX
KW   Carcinoembryonic antigen; domain III; domain A; domain B.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..89
FT                   /note="domain A"
FT   Domain          90..178
FT                   /note="domain B"
XX
PN   EP343946-A.
XX
PD   29-NOV-1989.
XX
PF   24-MAY-1989;   89EP-0305232.
XX
PR   25-MAY-1988;   88US-0198289.
XX
PA   (CITY ) CITY OF HOPE.
XX
PI   Shively JE;
XX
DR   WPI; 1989-349991/48.
DR   N-PSDB; N92449.
XX
PT   Carcinoembryonic antigen fragments - used in assays to determine the
PT   presence and amt. of the antigen in samples also contg. related antigens.
XX
PS   Disclosure; page 4; 15pp; English.
XX
CC   CEA fragments can be used in assays to determine the presence and amt. of
CC   CEA in samples which also may contain related antigens including its
CC   normal cross-reacting antigen or the 128 kD antigen.
XX
```

```
SQ   Sequence   178 AA;

  Query Match             84.4%;  Score 38;  DB 10;  Length 178;
  Best Local Similarity   77.8%;  Pred. No. 1.2;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
             |||||::||
Db     107 ylsganlnl 115


RESULT 12
R77436
ID   R77436 standard; Protein; 468 AA.
XX
AC   R77436;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-643) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..468
FT                   /note= "CEA (490-643)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 16; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   468 AA;
```

```
  Query Match             84.4%;  Score 38;  DB 16;  Length 468;
  Best Local Similarity   77.8%;  Pred. No. 3.8;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
           |||||::||
Db     396 ylsganlnl 404


RESULT  13
R77435
ID   R77435 standard; Protein; 493 AA.
XX
AC   R77435;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-C-terminal) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..493
FT                   /note= "CEA (490-C-terminal)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 15; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   493 AA;

  Query Match             84.4%;  Score 38;  DB 16;  Length 493;
  Best Local Similarity   77.8%;  Pred. No. 4;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
           |||||::||
Db     396 ylsganlnl 404


RESULT  14
R77437
ID   R77437 standard; Protein; 509 AA.
XX
AC   R77437;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-644)/BGP (391-430) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..469
FT                   /note= "CEA (490-644)"
FT   Protein         470..509
FT                   /note= "BGP (391-430)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 17; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   509 AA;

  Query Match             84.4%;  Score 38;  DB 16;  Length 509;
```

```
  Best Local Similarity   77.8%;  Pred. No. 4.2;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
           ||||||::||
Db     396 ylsganlnl 404


RESULT  15
R77438
ID   R77438 standard; Protein; 511 AA.
XX
AC   R77438;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-642)/BGP (387-430) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..467
FT                   /note= "CEA (490-642)"
FT   Protein         468..511
FT                   /note= "BGP (387-430)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 18; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   511 AA;


  Query Match             84.4%;  Score 38;  DB 16;  Length 511;
```

```
  Best Local Similarity   77.8%;  Pred. No. 4.2;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
           ||||||::||
Db     396 ylsganlnl 404


Search completed: December 16, 2000, 00:51:14
Job time: 18774 sec
```

```
                      GenCore version 4.5
           Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 04:22:12 ; Search time 111.26 Seconds
                                           (without alignments)
                                           7.553 Million cell updates/sec

Title:          US-09-529-121-3
Perfect score:  45
Sequence:       1 YLSGADINL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters:      436

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :      SPTREMBL_14:*
             1:  sp_archea:*
             2:  sp_bacteria:*
             3:  sp_fungi:*
             4:  sp_human:*
             5:  sp_invertebrate:*
             6:  sp_mammal:*
             7:  sp_mhc:*
             8:  sp_organelle:*
             9:  sp_phage:*
             10:  sp_plant:*
             11:  sp_rodent:*
             12:  sp_virus:*
             13:  sp_vertebrate:*
             14:  sp_unclassified:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                %
Result          Query
  No.   Score   Match Length DB  ID                   Description
--------------------------------------------------------------------
    1     16    35.6      8  5  Q94695               Q94695 physarum po
    2     16    35.6      9  1  Q50832               Q50832 methanococc
    3     15    33.3      7  11  Q63480               Q63480 rattus norv
    4     15    33.3      8  3  P87225               P87225 saccharomyc
    5     15    33.3      8  4  Q9UMC7               Q9umc7 homo sapien
    6     15    33.3      8  4  Q9UL56               Q9ul56 homo sapien
    7     14    31.1      8  2  Q9X3K1               Q9x3k1 prochloroco
    8     14    31.1      8  12  Q66807               Q66807 echovirus 2
    9     14    31.1      8  13  P82079               P82079 limnodynast
   10     14    31.1      8  13  Q9PS69               Q9ps69 gallus gall
   11     14    31.1      9  5  P82003               P82003 bombyx mori
   12     14    31.1      9  6  Q9TRS0               Q9trs0 oryctolagus
   13     14    31.1      9  11  O35953               O35953 mus musculu
   14     13    28.9      7  11  O55184               O55184 rattus norv
   15     13    28.9      9  2  Q44001               Q44001 aeromonas e
   16     13    28.9      9  2  Q44377               Q44377 aeromonas t
   17     13    28.9      9  2  Q44468               Q44468 aeromonas v
   18     13    28.9      9  2  Q43928               Q43928 aeromonas c
   19     13    28.9      9  2  Q9R7E8               Q9r7e8 escherichia
   20     12    26.7      7  12  Q07624               Q07624 rous sarcom
   21     12    26.7      8  4  Q15901               Q15901 homo sapien
   22     12    26.7      8  4  Q9Y4J3               Q9y4j3 homo sapien
   23     12    26.7      8  7  Q95213               Q95213 oryctolagus
   24     12    26.7      8  13  Q90498               Q90498 erythrura g
   25     12    26.7      8  13  Q91098               Q91098 manorina me
   26     12    26.7      9  2  Q9R635               Q9r635 chlamydia t
   27     12    26.7      9  4  Q16220               Q16220 homo sapien
   28     12    26.7      9  4  O95953               O95953 homo sapien
   29     12    26.7      9  5  O96417               O96417 drosophila
   30     12    26.7      9  5  Q27396               Q27396 babesia bov
   31     12    26.7      9  6  Q28112               Q28112 bos taurus
   32     11    24.4      7  12  Q9YQ10               Q9yq10 porcine tra
   33     11    24.4      8  2  Q51594               Q51594 escherichia
   34     11    24.4      8  2  Q9RQ57               Q9rq57 buchnera ap
   35     11    24.4      8  2  Q9RQ49               Q9rq49 buchnera ap
   36     11    24.4      8  4  Q15889               Q15889 homo sapien
   37     11    24.4      8  4  Q9UMH9               Q9umh9 homo sapien
   38     11    24.4      8  4  Q9UCN4               Q9ucn4 homo sapien
   39     11    24.4      8  6  Q9XSY1               Q9xsy1 canis famil
   40     11    24.4      8  8  Q35792               Q35792 saccharomyc
   41     11    24.4      8  11  Q9QVF4               Q9qvf4 rattus sp.
   42     11    24.4      8  12  Q83332               Q83332 murine hepa
   43     11    24.4      8  13  Q90493               Q90493 eopsaltria
   44     11    24.4      9  2  Q57328               Q57328 aeromonas s
   45     11    24.4      9  2  Q9R7H9               Q9r7h9 haemophilus
   46     11    24.4      9  2  Q9R5M1               Q9r5m1 staphylococ
   47     11    24.4      9  4  P78484               P78484 homo sapien
   48     11    24.4      9  5  Q9TWD6               Q9twd6 leptinotars
   49     11    24.4      9  6  Q9XSL0               Q9xsl0 capra hircu
   50     11    24.4      9  6  Q9TRW2               Q9trw2 oryctolagus
   51     11    24.4      9  12  Q89491               Q89491 murine minu
   52     11    24.4      9  13  Q92009               Q92009 gallus gall
   53     11    24.4      9  13  Q9PS68               Q9ps68 gallus gall
   54     10    22.2      8  2  Q47273               Q47273 escherichia
   55     10    22.2      8  2  P72221               P72221 pseudomonas
   56     10    22.2      8  2  O68485               O68485 klebsiella
   57     10    22.2      8  2  P77556               P77556 escherichia
   58     10    22.2      8  2  Q9R7T2               Q9r7t2 escherichia
   59     10    22.2      8  2  Q9R4M3               Q9r4m3 enterococcu
   60     10    22.2      8  3  Q9URB9               Q9urb9 saccharomyc
   61     10    22.2      8  4  Q16428               Q16428 homo sapien
   62     10    22.2      8  4  Q9Y4J4               Q9y4j4 homo sapien
   63     10    22.2      8  5  O02032               O02032 lytechinus
   64     10    22.2      8  5  Q94623               Q94623 manduca sex
   65     10    22.2      8  5  Q9VVJ2               Q9vvj2 drosophila
   66     10    22.2      8  5  Q9UB13               Q9ub13 albinaria h
   67     10    22.2      8  8  Q9XNP8               Q9xnp8 boophilus m
   68     10    22.2      8  8  Q9TKE5               Q9tke5 leptospermu
   69     10    22.2      8  8  Q9TD02               Q9td02 terranatos
   70     10    22.2      8  8  Q9T2W0               Q9t2w0 saccharomyc
   71     10    22.2      8  11  O35835               O35835 rattus norv
   72     10    22.2      8  13  P82082               P82082 limnodynast
   73     10    22.2      8  13  P82083               P82083 limnodynast
   74     10    22.2      9  2  Q51349               Q51349 pseudomonas
   75     10    22.2      9  2  Q9R9C4               Q9r9c4 borrelia bu

                                ALIGNMENTS

RESULT   1
Q94695
ID   Q94695      PRELIMINARY;     PRT;     8 AA.
AC   Q94695;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ACTIN (FRAGMENT).
GN   ARDC.
OS   Physarum polycephalum (Slime mold).
OC   Eukaryota; Myxogastria; Physarida; Physarum.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96182101.
RA   Benard M., Lagnel C., Pallotta D., Pierron G.;
RT   "Mapping of a replication origin within the promoter region of two
RT   unlinked, abundantly transcribed actin genes of Physarum
RT   polycephalum.";
RL   Mol. Cell. Biol. 16:968-976(1996).
DR   EMBL; M73459; AAB03706.1; -.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  878 MW;  F4C6C2CAAB187B16 CRC64;

  Query Match             35.6%;  Score 16;  DB 5;  Length 8;
  Best Local Similarity   33.3%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LSGADI 7
           : | |:
Db       1 MEGEDV 6

RESULT   2
Q50832
ID   Q50832      PRELIMINARY;      PRT;     9 AA.
AC   Q50832;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1996 (TrEMBLrel. 01, Last annotation update)
DE   INTERGENIC AT-RICH DNA SEQUENCE (FRAGMENT).
OS   Methanococcus voltae.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85230552.
RA   Bollschweiler C., Kuehn R., Klein A.;
RT   "Non-repetitive AT-rich sequences are found in intergenic regions of
RT   Methanococcus voltae DNA.";
RL   EMBO J. 4:805-809(1985).
DR   EMBL; X02518; CAA26355.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1087 MW;  99ED005DC404405A CRC64;

  Query Match             35.6%;  Score 16;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DIN 8
           |||
Db       2 DIN 4

RESULT   3
Q63480
ID   Q63480      PRELIMINARY;      PRT;     7 AA.
AC   Q63480;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   TR4-NS ORPHAN RECEPTOR (FRAGMENT).
GN   TR4.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96198747.
RA   Yoshikawa T., Makino S., Gao X.M., Xing G.Q., Chuang D.M.,
RA   Detera-Wadleigh S.D.;
RT   "Splice variants of rat TR4 orphan receptor: differential expression
RT   of novel sequences in the 5'-untranslated region and C-terminal
RT   domain.";
```

```
RL   Endocrinology 137:1562-1571(1996).
DR   EMBL; U59125; AAB02827.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   7 AA;  758 MW;  672AA87864005350 CRC64;

  Query Match             33.3%;  Score 15;  DB 11;  Length 7;
  Best Local Similarity   33.3%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LSGADI 7
           : | |:
Db       2 IRGGDL 7

RESULT   4
P87225
ID   P87225      PRELIMINARY;      PRT;     8 AA.
AC   P87225;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GIN11 PROTEIN (FRAGMENT).
GN   GIN11.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wedler H., Wedler E., Scharfe M., Wambutt R.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z73169; CAA97518.2; -.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  1019 MW;  4E21A9C449D5B73B CRC64;

  Query Match             33.3%;  Score 15;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLS 3
           |||
Db       1 YLS 3

RESULT   5
Q9UMC7
ID   Q9UMC7      PRELIMINARY;      PRT;     8 AA.
AC   Q9UMC7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   SHMT PROTEIN (FRAGMENT).
GN   SHMT.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chave K.J., Snell K., Sanders P.G.;
RT   "Isolation and characterisation of human genomic sequences encoding
RT   cytosolic serine hydroxymethyltransferase.";
RL   Biochem. Soc. Trans. 25:53-53(1997).
DR   EMBL; Y14492; CAB54844.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  868 MW;  7C205721E44AB5B8 CRC64;
```

```
  Query Match             33.3%;  Score 15;  DB 4;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 GADINL 9
            |:| :|
Db        1 GSDNHL 6


RESULT   6
Q9UL56
ID   Q9UL56      PRELIMINARY;      PRT;     8 AA.
AC   Q9UL56;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MUTANT NADH-CYTOCHROME B5 REDUCTASE (EC 1.6.2.2) (FRAGMENT).
GN   DIA1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fukumaki Y., Higasa K.;
RT   "Two novel mutations in Thai patients with hereditary
RT   methemoglobinemia types I and II: a subtle amino acid change causes
RT   instability of NADH-cytochrome b5 reductase.";
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF061830; AAF06818.1; -.
KW   Oxidoreductase.
FT   NON_TER       1      1
FT   VARIANT       9      9       -> R.
SQ   SEQUENCE   8 AA;  888 MW;  76C5B73B5051F6D8 CRC64;

  Query Match             33.3%;  Score 15;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLS 3
            |||
Db        5 YLS 7


RESULT   7
Q9X3K1
ID   Q9X3K1      PRELIMINARY;      PRT;     8 AA.
AC   Q9X3K1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CYTOCHROME B (FRAGMENT).
GN   PETB.
OS   Prochlorococcus sp.
OC   Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC   Prochlorococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Urbach E., Chisholm S.W.;
RT   "Genetic diversity in Prochlorococcus populations flow cytometrically
RT   sorted from the Sargasso Sea and Gulf Stream.";
RL   Limnol. Oceanog. 43:1615-1630(1998).
DR   EMBL; AF070193; AAD23233.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  799 MW;  10376865B72866D3 CRC64;

  Query Match             31.1%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSG 4
            |||
Db        4 LSG 6


RESULT   8
Q66807
ID   Q66807      PRELIMINARY;      PRT;     8 AA.
AC   Q66807;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   5'UTR IN IRES SEQUENCES (ISOLATE TH222) (FRAGMENT).
OS   Echovirus 25.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Enterovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TH222;
RA   Bailly J.L., Borman A:M., Peigue-Lafeuille H., Kean K.M.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X90724; CAA62259.1; -.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  821 MW;  EFC1B5A2D6DDD876 CRC64;

  Query Match             31.1%;  Score 14;  DB 12;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 GADIN 8
            || ::
Db        2 GAQVS 6


RESULT   9
P82079
ID   P82079      PRELIMINARY;      PRT;     8 AA.
AC   P82079;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   DYNASTIN 1.
OS   Limnodynastes interioris (Giant banjo frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Myobatrachidae;
OC   Limnodynastes.
RN   [1]
RP   SEQUENCE, AND MASS SPECTROMETRY.
RC   TISSUE=TIBIAL GLAND;
RA   Raftery M.J., Bradford A.M., Bowie J.H., Wallace J.C., Tyler M.J.;
RT   "Peptides from Australian frogs. The structure of the dynastins from
RT   the banjo frogs Limnodynastes interioras, Limnodynastes dumerilii and
RT   Limnodynastes terraereginae.";
RL   Aust. J. Chem. 46:833-842(1993).
CC   -!- MASS SPECTROMETRY: MW=729; METHOD=FAB.
KW   Amphibian skin.
SQ   SEQUENCE   8 AA;  729 MW;  7C28772865B72728 CRC64;

  Query Match             31.1%;  Score 14;  DB 13;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSG 4
            |||
Db        3 LSG 5


RESULT  10
Q9PS69
ID   Q9PS69      PRELIMINARY;      PRT;     8 AA.
AC   Q9PS69;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
```

```
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   LOW DENSITY LIPOPROTEIN RECEPTOR-RELATED PROTEIN.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92011685.
RA   Stifani S., Barber D.L., Aebersold R., Steyrer E., Shen X., Nimpf J.,
RA   Schneider W.J.;
RT   "The laying hen expresses two different low density lipoprotein
RT   receptor-related proteins.";
RL   J. Biol. Chem. 266:19079-19087(1991).
SQ   SEQUENCE   8 AA;  846 MW;  C007272DD865BAAA CRC64;

  Query Match             31.1%;  Score 14;  DB 13;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         3 SGA 5
             |||
Db         3 SGA 5

RESULT  11
P82003
ID   P82003      PRELIMINARY;      PRT;     9 AA.
AC   P82003;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   PROTHORACICOSTATIC PEPTIDE (PTSP).
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
RN   [1]
RP   SEQUENCE.
RC   STRAIN=C145 X N140; TISSUE=BRAIN;
RA   Hua Y.-J., Tanaka Y., Nakamura K.;
RT   "Identification of a prothoracicostatic peptide (PTSP) from the larval
RT   brain of the silkworm, Bombyx mori.";
RL   J. Biol. Chem. 0:0-0(1999).
CC   -!- FUNCTION: INHIBITS ECDYSTEROIDOGENESIS BY PROTHORACIC GLAND IN THE
CC       SILKWORM.
CC   -!- DEVELOPMENTAL STAGE: EARLY FIFTH INSTAR.
KW   Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1090 MW;  3878C5B4472AB6C3 CRC64;

  Query Match             31.1%;  Score 14;  DB 5;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         6 DIN 8
             |:|
Db         4 DLN 6

RESULT  12
Q9TRS0
ID   Q9TRS0      PRELIMINARY;      PRT;     9 AA.
AC   Q9TRS0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CALCYCLIN-ASSOCIATED PROTEIN, CAP50=CA2+/PHOSPHOLIPID-BINDING PROTEIN
DE   L-7 FRAGMENT.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92250478.
RA   Tokumitsu H., Mizutani A., Minami H., Kobayashi R., Hidaka H.;
RT   "A calcyclin-associated protein is a newly identified member of the
RT   Ca2+/phospholipid-binding proteins, annexin family.";
RL   J. Biol. Chem. 267:8919-8924(1992).
SQ   SEQUENCE   9 AA;  1010 MW;  64E419C44865B72B CRC64;

  Query Match             31.1%;  Score 14;  DB 6;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         2 LSG 4
             |||
Db         3 LSG 5

RESULT  13
O35953
ID   O35953      PRELIMINARY;      PRT;     9 AA.
AC   O35953;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   SODIUM CHANNEL, VOLTAGE-GATED, TYPE VIII, ALPHA POLYPEPTIDE
DE   (TRUNCATED VOLTAGE-GATED SODIUM CHANNEL ALPHA SUBUNIT) (FRAGMENT).
GN   SCN8A.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIII;
RX   MEDLINE; 97442476.
RA   Plummer N.W., McBurney M.W., Meisler M.H.;
RT   "Alternative splicing of the sodium channel SCN8A predicts a truncated
RT   two-domain protein in fetal brain and non-neuronal cells.";
RL   J. Biol. Chem. 272:24008-24015(1997).
DR   EMBL; U97672; AAB80914.1; -.
DR   MGD; MGI:103169; Scn8a.
KW   Ionic channel.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  898 MW;  22D92865B735B737 CRC64;

  Query Match             31.1%;  Score 14;  DB 11;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         2 LSG 4
             |||
Db         5 LSG 7

RESULT  14
O55184
ID   O55184      PRELIMINARY;      PRT;     7 AA.
AC   O55184;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ORPHAN RECEPTOR TR4-NS (FRAGMENT).
GN   TR4.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96198747.
RA   Yoshikawa T., Makino S., Gao X.M., Xing G.Q., Chuang D.M.,
```

```
RA   Detera-Wadleigh S.D.;
RT   "Splice variants of rat TR4 orphan receptor: differential expression
RT   of novel sequences in the 5'-untranslated region and C-terminal
RT   domain.";
RL   Endocrinology 137:1562-1571(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96299786.
RA   Yoshikawa T., Dupont B.R., Leach R.J., Detera-Wadleigh S.D.;
RT   "New variants of the human and rat nuclear hormone receptor, TR4:
RT   expression and chromosomal localization of the human gene.";
RL   Genomics 35:361-366(1996).
DR   EMBL; U59454; AAB91433.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   7 AA;  663 MW;  6DDAA8787EB05350 CRC64;


  Query Match             28.9%;  Score 13;  DB 11;  Length 7;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LSGAD 6
           : | |
Db       2 ICGGD 6

RESULT  15
Q44001
ID   Q44001      PRELIMINARY;      PRT;     9 AA.
AC   Q44001;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (STRAIN CDC 9179-79) (FRAGMENT).
GN   EXEF'.
OS   Aeromonas eucrenophila.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 9179-79;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89461; CAA61637.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1059 MW;  785852C69444472B CRC64;

  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 INL 9
           :||
Db       7 VNL 9

RESULT  16
Q44377
ID   Q44377      PRELIMINARY;      PRT;     9 AA.
AC   Q44377;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (STRAIN ATCC 49659) (FRAGMENT).
GN   EXEF'.
OS   Aeromonas trota.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 49659;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89468; CAA61651.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1059 MW;  785852C69444472B CRC64;

  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 INL 9
           :||
Db       7 VNL 9

RESULT  17
Q44468
ID   Q44468      PRELIMINARY;      PRT;     9 AA.
AC   Q44468;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (STRAIN CDC 1306-83) (FRAGMENT).
GN   EXEF'.
OS   Aeromonas veronii.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1306-83;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89457; CAA61629.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1059 MW;  785852C69444472B CRC64;

  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 INL 9
           :||
Db       7 VNL 9

RESULT  18
Q43928
ID   Q43928      PRELIMINARY;      PRT;     9 AA.
AC   Q43928; Q43918; Q43920; Q43921;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (FRAGMENT).
GN   EXEF'.
OS   Aeromonas caviae.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VARIOUS STRAINS;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
```

```
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89464; CAA61643.1; -.
DR   EMBL; X89462; CAA61639.1; -.
DR   EMBL; X89460; CAA61635.1; -.
DR   EMBL; X89463; CAA61641.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1059 MW;  785852C69444472B CRC64;

  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        7 INL 9
            :||
Db        7 VNL 9

RESULT  19
Q9R7E8
ID   Q9R7E8      PRELIMINARY;      PRT;     9 AA.
AC   Q9R7E8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   KPSD PROTEIN (FRAGMENT).
GN   KPSD.
OS   Escherichia coli.
OG   Plasmid pCR3.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95180691.
RA   Rosenow C., Roberts I.S., Jann K.;
RT   "Isolation from recombinant Escherichia coli and characterization of
RT   CMP-Kdo synthetase, involved in the expression of the capsular K5
RT   polysaccharide (K-CKS).";
RL   FEMS Microbiol. Lett. 125:159-164(1995).
DR   EMBL; S76943; CAB33515.1; -.
KW   Plasmid.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  899 MW;  3EBBB72042C33DD8 CRC64;

  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        4 GADINL 9
            || : |
Db        2 GAKVIL 7

RESULT  20
Q07624
ID   Q07624      PRELIMINARY;      PRT;     7 AA.
AC   Q07624;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL PROTEIN (FRAGMENT).
OS   Rous sarcoma virus.
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PRAGUE C;
RX   MEDLINE; 93010967.
RA   Donze O., Spahr P.F.;
RT   "Role of the open reading frames of Rous sarcoma virus leader RNA in
RT   translation and genome packaging.";
RL   EMBO J. 11:3747-3757(1992).
```

```
DR   EMBL; X67587; CAA47862.1; -.
KW   Hypothetical protein.
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  672 MW;  776045A7687DD6F0 CRC64;

  Query Match             26.7%;  Score 12;  DB 12;  Length 7;
  Best Local Similarity   33.3%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        2 LSGADI 7
            ::|  |
Db        1 MAGPSI 6

RESULT  21
Q15901
ID   Q15901      PRELIMINARY;      PRT;     8 AA.
AC   Q15901;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   (CLONE XP7B11B) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   Lee C.C., Yazdani A., Wehnert M., Bailey J., Couch L., Xiong M.,
RA   Coolbaugh M.I., Chinault C.A., Baldini A., Lindsay E.A., Zhao Z.Y.,
RA   Caskey C.T.H.;
RL   Hum. Mol. Genet. 0:0-0(0).
DR   EMBL; L32080; AAA73891.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  860 MW;  37D72878676729CB CRC64;

  Query Match             26.7%;  Score 12;  DB 4;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSG 4
            :| |
Db        2 FLPG 5

RESULT  22
Q9Y4J3
ID   Q9Y4J3      PRELIMINARY;      PRT;     8 AA.
AC   Q9Y4J3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   RUNT/68NT/MTG8 PROTEIN (FRAGMENT).
GN   RUNT/68NT/MTG8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE MARROW;
RX   MEDLINE; 95002916.
RA   Tighe J.E., Calabi F.;
RT   "Alternative, out-of-frame runt/MTG8 transcripts are encoded by the
RT   derivative (8) chromosome in the t(8;21) of acute myeloid leukemia
RT   M2.";
RL   Blood 84:2115-2121(1994).
DR   EMBL; S74094; AAD14973.2; -.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  929 MW;  30B764405B17244B CRC64;
```

```
  Query Match             26.7%;  Score 12;  DB 4;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DIN 8
           :||
Db       4 EIN 6


RESULT  23
Q95213
ID   Q95213      PRELIMINARY;      PRT;     8 AA.
AC   Q95213;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   GERMLINE DH (DF) GENE (FRAGMENT).
GN   DF.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F-I/RGM;
RA   Mage R.G., Chen H.T., Alexander C.B., Chen F.F.;
RL   Mol. Immunol. 0:0-0(0).
DR   EMBL; U62585; AAB18735.1; -.
KW   MHC.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  845 MW;  5CA861B5AB58677B CRC64;


  Query Match             26.7%;  Score 12;  DB 7;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           | :|
Db       4 YSTG 7


RESULT  24
Q90498
ID   Q90498      PRELIMINARY;      PRT;     8 AA.
AC   Q90498;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   MYOGLOBIN (FRAGMENT).
OS   Erythrura gouldiae (Gouldian finch).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Passeridae; Erythrura.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EGG1;
RX   MEDLINE; 98208049.
RA   Heslewood M.M., Elphinstone M.S., Tidemann S.C., Baverstock P.R.;
RT   "Myoglobin intron variation in the Gouldian Finch Erythrura gouldiae
RT   assessed by temperature gradient gel electrophoresis.";
RL   Electrophoresis 19:142-151(1998).
DR   EMBL; U40496; AAC60363.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  890 MW;  C6C1F2C865B046DE CRC64;


  Query Match             26.7%;  Score 12;  DB 13;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       3 ISG 5


RESULT  25
Q91098
ID   Q91098      PRELIMINARY;      PRT;     8 AA.
AC   Q91098;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   MYOGLOBIN (FRAGMENT).
OS   Manorina melanocephala (noisy miner).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Meliphagidae; Manorina.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D02;
RX   MEDLINE; 98208049.
RA   Heslewood M.M., Elphinstone M.S., Tidemann S.C., Baverstock P.R.;
RT   "Myoglobin intron variation in the Gouldian Finch Erythrura gouldiae
RT   assessed by temperature gradient gel electrophoresis.";
RL   Electrophoresis 19:142-151(1998).
DR   EMBL; U40497; AAC60364.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  890 MW;  C6C1F2C865B046DE CRC64;


  Query Match             26.7%;  Score 12;  DB 13;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       3 ISG 5


RESULT  26
Q9R635
ID   Q9R635      PRELIMINARY;      PRT;     9 AA.
AC   Q9R635;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR OUTER MEMBRANE PROTEIN VARIABLE DOMAIN IV, MOMP VD IV.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92040090.
RA   Peterson E.M., Cheng X., Markoff B.A., Fielder T.J., de la Maza L.M.;
RT   "Functional and structural mapping of Chlamydia trachomatis species-
RT   specific major outer membrane protein epitopes by use of neutralizing
RT   monoclonal antibodies.";
RL   Infect. Immun. 59:4147-4153(1991).
SQ   SEQUENCE   9 AA;  976 MW;  9C61B041B7645361 CRC64;


  Query Match             26.7%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       7 ISG 9


RESULT  27
Q16220
ID   Q16220      PRELIMINARY;      PRT;     9 AA.
AC   Q16220;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
```

```
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HGRP PROTEIN (FRAGMENT).
GN   HGRP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94320083.
RA   Nagalla S.R., Spindel E.R.;
RT   "Functional analysis of the 5'-flanking region of the human gastrin-
RT   releasing peptide gene in small cell lung carcinoma cell lines.";
RL   Cancer Res. 54:4461-4467(1994).
DR   EMBL; S73265; AAD14116.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1070 MW;  77FE37672B040864 CRC64;

  Query Match              26.7%;  Score 12;  DB 4;  Length 9;
  Best Local Similarity   25.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 LSGADINL 9
              : | :: |
Db          1 MRGRELPL 8


RESULT  28
O95953
ID   O95953      PRELIMINARY;      PRT;     9 AA.
AC   O95953;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   GALACTOCEREBROSIDASE (EC 3.2.1.46) (FRAGMENT).
GN   GALC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Lulli L., Torchiana E., Finocchiaro G.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U77631; AAD15626.1; -.
KW   Hydrolase; Glycosidase.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1069 MW;  374E2AADC2C699C8 CRC64;

  Query Match              26.7%;  Score 12;  DB 4;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 ADI 7
              ||:
Db          6 ADL 8


RESULT  29
O96417
ID   O96417      PRELIMINARY;      PRT;     9 AA.
AC   O96417;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   SXL E1 FORM (FRAGMENT).
GN   SXL.
OS   Drosophila virilis (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98337843.
RA   Erickson J.W., Cline T.W.;
RT   "Key aspects of the primary sex determination mechanism are conserved
RT   across the genus Drosophila.";
RL   Development 125:3259-3268(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Bell M., Cline T.W.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF046045; AAC97605.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1089 MW;  9A8BD1AAA9C449CA CRC64;

  Query Match              26.7%;  Score 12;  DB 5;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          6 DIN 8
              | |
Db          2 DFN 4


RESULT  30
Q27396
ID   Q27396      PRELIMINARY;      PRT;     9 AA.
AC   Q27396;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   RHOPTRY ASSOCIATED PROTEIN 1.
GN   RAP-1.
OS   Babesia bovis.
OC   Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MO7;
RA   Suarez C.E., Palmer G.H., Hines S.A., McElwain T.F.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L77326; AAA96415.1; -.
SQ   SEQUENCE   9 AA;  931 MW;  6D5BD2C865B05044 CRC64;

  Query Match              26.7%;  Score 12;  DB 5;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LSG 4
              :||
Db          4 ISG 6


RESULT  31
Q28112
ID   Q28112      PRELIMINARY;      PRT;     9 AA.
AC   Q28112;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   6-PHOSPHOFRUCTO-2-KINASE/FRUCTOSE-2, 6-BISPHOSPHATASE (FRAGMENT).
GN   GENE B.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93387464.
RA   Vidal H., Crepin K.M., Rider M.H., Hue L., Rousseau G.G.;
RT   "Cloning and expression of novel isoforms of 6-phosphofructo-2-
RT   kinase/fructose-2,6-bisphosphatase from bovine heart.";
```

```
RL   FEBS Lett. 330:329-333(1993).
DR   EMBL; X74564; CAA52652.1; -.
FT   NON_TER      9      9
SQ   SEQUENCE   9 AA;  837 MW;  859CA5BDC7644865 CRC64;


  Query Match             26.7%;  Score 12;  DB 6;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       1 MSG 3


RESULT  32
Q9YQ10
ID   Q9YQ10      PRELIMINARY;      PRT;     7 AA.
AC   Q9YQ10;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HYPOTHETICAL FUSION PROTEIN.
OS   porcine transmissible gastroenteritis virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99099045.
RA   Izeta A., Smerdou C., Alonso S., Penzes Z., Mendez A., Plana-Duran J.,
RA   Enjuanes L.;
RT   "Replication and packaging of transmissible gastroenteritis
RT   coronavirus-derived synthetic minigenomes.";
RL   J. Virol. 73:1535-1545(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95159435.
RA   Eleouet J., Rasschaert D., Lambert P., Levy L., Vende P., Laude H.;
RT   "Complete sequence (20 kilobases) of the polyprotein-encoding gene 1
RT   of transmissible gastroenteritis virus.";
RL   Virology 206:817-822(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88078100.
RA   Rasschaert D., Gelfi J., Laude H.;
RT   "Enteric coronavirus TGEV: partial sequence of the genomic RNA, its
RT   organization and expression.";
RL   Biochimie 69:591-600(1987).
DR   EMBL; AJ011482; CAA09625.1; -.
SQ   SEQUENCE   7 AA;  927 MW;  69D6D7273B5726F0 CRC64;


  Query Match             24.4%;  Score 11;  DB 12;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YL 2
           ||
Db       3 YL 4


RESULT  33
Q51594
ID   Q51594      PRELIMINARY;      PRT;     8 AA.
AC   Q51594;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   COPB PROTEIN (FRAGMENT).
OS   Escherichia coli.
OG   Plasmid IncFI ColV2-K94.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86223772.
RA   Weber P.C., Palchaudhuri S.;
RT   "Incompatibility repressor in a RepA-like replicon of the IncFI
RT   plasmid ColV2-K94.";
RL   J. Bacteriol. 166:1106-1112(1986).
DR   EMBL; M13472; AAA23194.1; -.
KW   Plasmid.
FT   NON_TER      1      1
SQ   SEQUENCE   8 AA;  998 MW;  0F37205AA73416D7 CRC64;


  Query Match             24.4%;  Score 11;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 LSGADI 7
           |   ||
Db       1 LQRLDI 6


RESULT  34
Q9RQ57
ID   Q9RQ57      PRELIMINARY;      PRT;     8 AA.
AC   Q9RQ57;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   NIFS PROTEIN HOMOLOG (FRAGMENT).
GN   NIFS.
OS   Buchnera aphidicola.
OC   Bacteria; Proteobacteria; gamma subdivision; Buchnera.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 20022990.
RA   Clark M.A., Moran N.A., Baumann P.;
RT   "Sequence evolution in bacterial endosymbionts having extreme base
RT   compositions.";
RL   Mol. Biol. Evol. 16:1586-1598(1999).
DR   EMBL; AF130812; AAF13797.1; -.
FT   NON_TER      8      8
SQ   SEQUENCE   8 AA;  980 MW;  F3A73B504771A336 CRC64;


  Query Match             24.4%;  Score 11;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YL 2
           ||
Db       6 YL 7


RESULT  35
Q9RQ49
ID   Q9RQ49      PRELIMINARY;      PRT;     8 AA.
AC   Q9RQ49;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   NIFS PROTEIN HOMOLOG (FRAGMENT).
GN   NIFS.
OS   Buchnera aphidicola.
OC   Bacteria; Proteobacteria; gamma subdivision; Buchnera.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 20022990.
RA   Clark M.A., Moran N.A., Baumann P.;
RT   "Sequence evolution in bacterial endosymbionts having extreme base
RT   compositions.";
RL   Mol. Biol. Evol. 16:1586-1598(1999).
DR   EMBL; AF130814; AAF13805.1; -.
```

```
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  992 MW;  F3A73B5047672336 CRC64;

  Query Match             24.4%;  Score 11;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YL 2
            ||
Db        6 YL 7


Search completed: December 16, 2000, 04:22:13
Job time: 4607 sec
```

```
                    GenCore version 4.5
           Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 04:23:31 ; Search time 68.03 Seconds
                                               (without alignments)
                                               4.227 Million cell updates/sec

Title:           US-09-529-121-3
Perfect score:   45
Sequence:        1 YLSGADINL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters:      213

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       SwissProt_39:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                               SUMMARIES
                %
Result         Query
  No.   Score  Match Length DB  ID                       Description
-------------------------------------------------------------------------
    1      17   37.8      8  1  CPD1_ENTFA               P13269 enterococcu
    2      16   35.6      8  1  LCK8_LEUMA               P19990 leucophaea
    3      16   35.6      9  1  ISOT_CYPCA               P42993 cyprinus ca
    4      15   33.3      8  1  PLP_BRANA                P81707 brassica na
    5      15   33.3      9  1  FAR6_CALVO               P41861 calliphora
    6      14   31.1      8  1  ALL5_CYDPO               P82156 cydia pomon
    7      14   31.1      9  1  LMIP_LOCMI               P31799 locusta mig
    8      13   28.9      7  1  LANC_CARUI               P36960 carnobacter
    9      13   28.9      9  1  DSIP_RABIT               P01158 oryctolagus
   10      13   28.9      9  1  FAR2_PANRE               P41873 panagrellus
   11      13   28.9      9  1  OXYT_RAJCL               P42994 raja clavat
   12      12   26.7      8  1  ACT_CARMA                P80709 carcinus ma
   13      12   26.7      8  1  CCKN_MACEU               P30369 macropus eu
   14      12   26.7      9  1  FARD_CALVO               P41868 calliphora
   15      12   26.7      9  1  OXYA_SCYCA               P42996 scyliorhinu
   16      12   26.7      9  1  OXYT_BUFRE               P42995 bufo regula
   17      12   26.7      9  1  PGLR_DIAAB               P81179 diaprepes a
   18      12   26.7      9  1  UPA6_HUMAN               P30092 homo sapien
   19      11   24.4      4  1  FAR3_HIRME               P42562 hirudo medi
   20      11   24.4      5  1  PRCT_PERAM               P01373 periplaneta
   21      11   24.4      7  1  FAR2_ASCSU               P31890 ascaris suu
   22      11   24.4      7  1  GFRP_MOUSE               P99025 mus musculu
   23      11   24.4      8  1  CAD1_ENTFA               P13268 enterococcu
   24      11   24.4      8  1  FAR8_CALVO               P41863 calliphora
   25      11   24.4      9  1  D1_NEPNO                 P24816 nephrops no
   26      11   24.4      9  1  FAR5_CALVO               P41860 calliphora
   27      11   24.4      9  1  FAR7_CALVO               P41862 calliphora
   28      11   24.4      9  1  MOSF_CLYJA               P19853 clypeaster
   29      11   24.4      9  1  OXYA_SQUAC               P42999 squalus aca
   30      11   24.4      9  1  OXYV_SQUAC               P43000 squalus aca
   31      10   22.2      4  1  ACH1_ACHFU               P35904 achatina fu
   32      10   22.2      5  1  UXA4_CHLTR               P38005 chlamydia t
   33      10   22.2      6  1  CIP2_MYTED               P13737 mytilus edu
```

```
   34      10   22.2      6  1  TMOF_SARBU               P41495 sarcophaga
   35      10   22.2      6  1  UN06_CLOPA               P81351 clostridium
   36      10   22.2      7  1  ALL2_CARMA               P81805 carcinus ma
   37      10   22.2      7  1  ALL3_CARMA               P81806 carcinus ma
   38      10   22.2      7  1  ALL4_CARMA               P81807 carcinus ma
   39      10   22.2      7  1  ALL5_CARMA               P81808 carcinus ma
   40      10   22.2      7  1  ALL7_CYDPO               P82158 cydia pomon
   41      10   22.2      7  1  UN06_PINPS               P81675 pinus pinas
   42      10   22.2      8  1  AL12_CARMA               P81815 carcinus ma
   43      10   22.2      8  1  AL17_CARMA               P81820 carcinus ma
   44      10   22.2      8  1  ALL7_CARMA               P81809 carcinus ma
   45      10   22.2      8  1  ALL8_CARMA               P81811 carcinus ma
   46      10   22.2      8  1  ALL9_CARMA               P81812 carcinus ma
   47      10   22.2      8  1  FAR1_PANRE               P41872 panagrellus
   48      10   22.2      8  1  FAR3_HOMAM               P41486 homarus ame
   49      10   22.2      8  1  GLUR_HUMAN               P02729 homo sapien
   50      10   22.2      8  1  LCK5_LEUMA               P19987 leucophaea
   51      10   22.2      8  1  UF06_MOUSE               P38644 mus musculu
   52      10   22.2      8  1  VGLG_HSV2B               P81780 herpes simp
   53      10   22.2      9  1  AL10_CARMA               P81813 carcinus ma
   54      10   22.2      9  1  AL11_CARMA               P81814 carcinus ma
   55      10   22.2      9  1  BUK_CLOPA                P81337 clostridium
   56      10   22.2      9  1  FAR5_ASCSU               P43170 ascaris suu
   57      10   22.2      9  1  FIBB_MACFU               P19345 macaca fusc
   58      10   22.2      9  1  MOSH_CLYJA               P19852 clypeaster
   59      10   22.2      9  1  OXYF_SCYCA               P42997 scyliorhinu
   60      10   22.2      9  1  OXYT_RABIT               P32878 oryctolagus
   61      10   22.2      9  1  TKL1_LOCMI               P16223 locusta mig
   62      10   22.2      9  1  TRP4_LEUMA               P81736 leucophaea
   63      10   22.2      9  1  ULAH_HUMAN               P31934 homo sapien
   64       9   20.0      4  1  EOSI_HUMAN               P02731 homo sapien
   65       9   20.0      4  1  FAR4_HIRME               P42563 hirudo medi
   66       9   20.0      5  1  AL14_CARMA               P81817 carcinus ma
   67       9   20.0      5  1  FARP_ARTTR               P41853 artioposthi
   68       9   20.0      7  1  FAR2_PROCL               P38498 procambarus
   69       9   20.0      7  1  FAR3_HAECO               P81298 haemonchus
   70       9   20.0      7  1  FAR3_PANRE               P41874 panagrellus
   71       9   20.0      7  1  FAR5_HIRME               P42564 hirudo medi
   72       9   20.0      7  1  UF03_MOUSE               P38641 mus musculu
   73       9   20.0      8  1  AKHG_GRYBI               P14086 gryllus bim
   74       9   20.0      8  1  AKH_LIBAU                P25418 libellula a
   75       9   20.0      8  1  AL15_CARMA               P81818 carcinus ma


                               ALIGNMENTS


RESULT   1
CPD1_ENTFA
ID   CPD1_ENTFA     STANDARD;      PRT;     8 AA.
AC   P13269;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   SEX PHEROMONE CPD1.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 85040388.
RA   Suzuki A., Mori M., Sagakami Y., Isogai A., Fujino M., Kitada C.,
RA   Craig R.A., Clewell D.B.;
RT   "Isolation and structure of bacterial sex pheromone, cPD1.";
RL   Science 226:849-850(1984).
CC   -!- FUNCTION: CPD1 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE
CC       BACTERIOCIN PLASMID PPD1.
KW   Pheromone.
SQ   SEQUENCE   8 AA;  913 MW;  8665B729C682C729 CRC64;

  Query Match             37.8%;  Score 17;  DB 1;  Length 8;
  Best Local Similarity   75.0%;  Pred. No. 8.8e+04;
```

```
  Matches     3;  Conservative     1;  Mismatches     0;  Indels     0;  Gaps     0;

Qy       1 YLSG 4
           :|||
Db       5 FLSG 8

RESULT   2
LCK8_LEUMA
ID   LCK8_LEUMA     STANDARD;      PRT;     8 AA.
AC   P19990;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   LEUCOKININ VIII (L-VIII).
OS   Leucophaea maderae (Madeira cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blaberoidea; Blaberidae; Leucophaea.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=HEAD;
RA   Holman G.M., Cook B.J., Nachman R.J.;
RT   "Isolation, primary structure and synthesis of leucokinins VII and
RT   VIII: the final members of this new family of cephalomyotropic
RT   peptides isolated from head extracts of Leucophaea maderae.";
RL   Comp. Biochem. Physiol. 88C:31-34(1987).
CC   -!- FUNCTION: THIS CEPHALOMYOTROPIC PEPTIDE STIMULATES CONTRACTILE
CC       ACTIVITY OF COCKROACH PROTODEUM (HINDGUT).
CC   -!- SIMILARITY: TO THE OTHER LEUCOKININS.
DR   PIR; JS0318; JS0318.
KW   Neuropeptide; Amidation.
FT   MOD_RES      8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  902 MW;  736365AB59CAADD8 CRC64;


  Query Match             35.6%;  Score 16;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches     3;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy       4 GAD 6
           |||
Db       1 GAD 3

RESULT   3
ISOT_CYPCA
ID   ISOT_CYPCA     STANDARD;      PRT;     9 AA.
AC   P42993;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ISOTOCIN.
OS   Cyprinus carpio (Common carp).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PITUITARY;
RA   Acher R., Chauvet J., Chauvet M.-T., Crepy D.;
RT   "Characterization of neurohypophyseal hormones from a fresh water bony
RT   fish, the carp (Cyprinus carpio). Comparison with hormones from sea
RT   water bony fishs.";
RL   Comp. Biochem. Physiol. 14:245-254(1965).
CC   -!- FUNCTION: ANTIDIURETIC HORMONE.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   PIR; A61364; A61364.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
```

```
FT   MOD_RES      9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  969 MW;  17FF476EB455B04B CRC64;

  Query Match             35.6%;  Score 16;  DB 1;  Length 9;
  Best Local Similarity   42.9%;  Pred. No. 8.8e+04;
  Matches     3;  Conservative     1;  Mismatches     3;  Indels     0;  Gaps     0;

Qy       1 YLSGADI 7
           |:|   |
Db       2 YISNCPI 8

RESULT   4
PLP_BRANA
ID   PLP_BRANA      STANDARD;      PRT;     8 AA.
AC   P81707;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PLASTIDIAL LIPID-ASSOCIATED PROTEIN (FRAGMENT).
OS   Brassica napus (Rape).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Brassica.
RN   [1]
RP   SEQUENCE.
RC   STRAIN=CV. TOPAZ; TISSUE=TAPETUM;
RX   MEDLINE; 99349136.
RA   Hernandez-Pinzon I., Ross J.H.E., Barnes K.A., Damant A.P.,
RA   Murphy D.J.;
RT   "Composition and role of tapetal lipid bodies in the biogenesis of the
RT   pollen coat of Brassica napus.";
RL   Planta 208:588-598(1999).
CC   -!- FUNCTION: MAY PLAY A STRUCTURAL ROLE IN THE ELAIOPLAST, A TAPETUM-
CC       SPECIFIC PLASTIDIAL LIPID ORGANELLE.
CC   -!- TISSUE SPECIFICITY: TAPETUM OF ANTHERS.
FT   NON_TER      8      8
SQ   SEQUENCE   8 AA;  989 MW;  9D7B1AA452CAA042 CRC64;


  Query Match             33.3%;  Score 15;  DB 1;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches     2;  Conservative     1;  Mismatches     0;  Indels     0;  Gaps     0;

Qy       6 DIN 8
           |:|
Db       3 DVN 5


RESULT   5
FAR6_CALVO
ID   FAR6_CALVO     STANDARD;      PRT;     9 AA.
AC   P41861;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 6.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
```

```
CC       FAMILY.
DR   PIR; F41978; F41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1058 MW;  96D10699CAB6D865 CRC64;


  Query Match             33.3%;  Score 15;  DB 1;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 8.8e+04;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        3 SGAD 6
            || |
Db        2 SGQD 5


RESULT   6
ALL5_CYDPO
ID   ALL5_CYDPO     STANDARD;      PRT;     8 AA.
AC   P82156;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYDIASTATIN 5.
OS   Cydia pomonella (Codling moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Tortricoidea; Tortricidae; Olethreutinae; Cydia.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LARVA;
RX   MEDLINE; 98054539.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
RA   Davey M., East P.D., Thorpe A.;
RT   "Lepidopteran peptides of the allatostatin superfamily.";
RL   Peptides 18:1301-1309(1997).
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  898 MW;  922879CABB58640D CRC64;


  Query Match             31.1%;  Score 14;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    3;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        4 GADINL 9
            | |  |
Db        3 GYDFGL 8


RESULT   7
LMIP_LOCMI
ID   LMIP_LOCMI     STANDARD;      PRT;     9 AA.
AC   P31799;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-OCT-1993 (Rel. 27, Last annotation update)
DE   LOCUSTAMYOINHIBITING PEPTIDE (LOM-MIP).
OS   Locusta migratoria (Migratory locust).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC   Acridomorpha; Acridoidea; Acrididae; Locusta.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92179466.
RA   Schoofs L., Holman G.M., Hayes T.K., Nachman R.J., de Loof A.;
RT   "Isolation, identification and synthesis of locustamyoinhibiting
RT   peptide (LOM-MIP), a novel biologically active neuropeptide from
RT   Locusta migratoria.";
RL   Regul. Pept. 36:111-119(1991).
CC   -!- FUNCTION: SUPPRESSES SPONTANEOUS CONTRACTIONS OF THE HINDGUT AND
CC       OVIDUCT.
```

```
CC   -!- TISSUE SPECIFICITY: NEURONS LOCATED IN TWO VENTRAL CELL CLUSTERS
CC       IN THE SUBOESOPHAGEAL GANGLION.
DR   PIR; A60065; AKLQIM.
KW   Amidation; Neuropeptide.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1060 MW;  387D7DD4472AB6C3 CRC64;


  Query Match             31.1%;  Score 14;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 DIN 8
            |:|
Db        4 DLN 6


RESULT   8
LANC_CARUI
ID   LANC_CARUI     STANDARD;      PRT;     7 AA.
AC   P36960;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   LANTIBIOTIC CARNOCIN UI49 (FRAGMENT).
OS   Carnobacterium sp. (strain UI49).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC   Carnobacterium.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92321768.
RA   Stoffels G., Nissen-Meyer J., Gudmundsdottir A., Sletten K., Holo H.,
RA   Nes I.F.;
RT   "Purification and characterization of a new bacteriocin isolated from
RT   a Carnobacterium sp.";
RL   Appl. Environ. Microbiol. 58:1417-1422(1992).
CC   -!- FUNCTION: LANTHIONINE-CONTAINING PEPTIDE ANTIBIOTIC (LANTIBIOTIC).
CC       ACTIVE ON GRAM-POSITIVE BACTERIA.
KW   Antibiotic; Lantibiotic.
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  786 MW;  741776D05B05B810 CRC64;


  Query Match             28.9%;  Score 13;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        4 GADI 7
            |::|
Db        1 GSEI 4


RESULT   9
DSIP_RABIT
ID   DSIP_RABIT     STANDARD;      PRT;     9 AA.
AC   P01158;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DELTA SLEEP-INDUCING PEPTIDE (DSIP).
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 77185324.
RA   Monnier M., Dudler L., Gachter R., Maier P.F., Tobler H.J.,
RA   Schoenenberger G.A.;
RT   "The delta sleep inducing peptide (DSIP). Comparative properties of
RT   the original and synthetic nonapeptide.";
RL   Experientia 33:548-552(1977).
RN   [2]
RP   SEQUENCE, AND SYNTHESIS.
```

```
RX   MEDLINE; 79054421.
RA   Schoenenberger G.A., Maier P.F., Tobler H.J., Wilson K., Monnier M.;
RT   "The delta EEG (sleep)-inducing peptide (DSIP). XI. Amino-acid
RT   analysis, sequence, synthesis and activity of the nonapeptide.";
RL   Pflugers Arch. 376:119-129(1978).
CC   -!- FUNCTION: WHEN INFUSED INTO THE MESODIENCEPHALIC VENTRICLE OF
CC       RECIPIENT RABBITS INDUCES SPINDLE AND DELTA EEG ACTIVITY AND
CC       REDUCED MOTOR ACTIVITIES.
CC   -!- MISCELLANEOUS: THIS PEPTIDE WAS OBTAINED FROM DIALYSATES OF
CC       OCCIPITAL VENOUS SINUS BLOOD FROM RABBITS KEPT ASLEEP BY ELECTRIC
CC       STIMULATION OF THE THALAMUS.
DR   PIR; A01422; QDRB.
SQ   SEQUENCE   9 AA;  849 MW;  DDD365BDDAA8787D CRC64;

  Query Match             28.9%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        3 SGAD 6
            :| |
Db        2 AGGD 5


RESULT  10
FAR2_PANRE
ID   FAR2_PANRE     STANDARD;      PRT;     9 AA.
AC   P41873;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE PF2 (SADPNFLRF-AMIDE).
OS   Panagrellus redivivus.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
OC   Panagrolaimoidea; Panagrolaimidae; Panagrellus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 93027659.
RA   Geary T.G., Price D.A., Bowman J.W., Winterrowd C.A., Mackenzie C.D.,
RA   Garrison R.D., Williams J.F., Friedman A.R.;
RT   "Two FMRFamide-like peptides from the free-living nematode
RT   Panagrellus redivivus.";
RL   Peptides 13:209-214(1992).
CC   -!- FUNCTION: MYOACTIVE.
CC   -!- TISSUE SPECIFICITY: NERVE CORDS AND PAIRED GROUPS OF CELLS LOCATED
CC       CAUDALLY TO THE BASE OF THE PHARYNX.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1066 MW;  DA0B0729C4576AAD CRC64;

  Query Match             28.9%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 8.8e+04;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        5 ADIN 8
            || |
Db        2 ADPN 5


RESULT  11
OXYT_RAJCL
ID   OXYT_RAJCL     STANDARD;      PRT;     9 AA.
AC   P42994;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   GLUMITOCIN.
OS   Raja clavata (Thornback ray).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Hypnosqualea; Pristiorajea; Batoidea;
OC   Rajiformes; Rajidae; Raja.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 66123415.
RA   Acher R., Chauvet J., Chauvet M.-T., Crepy D.;
RT   "Phylogeny of neurophyophyseal peptides: isolation of a new hormone,
RT   glumitocin (Ser 4-Gln 8-ocytocin) present in a cartilaginous fish,
RT   the ray (Raia clavata).";
RL   Biochim. Biophys. Acta 107:393-396(1965).
CC   -!- FUNCTION: ANTIDIURETIC HORMONE.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  984 MW;  17E9C76EB455B04B CRC64;

  Query Match             28.9%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLS 3
            |:|
Db        2 YIS 4


RESULT  12
ACT_CARMA
ID   ACT_CARMA      STANDARD;      PRT;     8 AA.
AC   P80709;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   ACTIN (FRAGMENT).
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RA   Lachaise F., Somme G., Carpentier G., Granjeon E., Webster S.,
RA   Baghdassarian D.;
RT   "A transaldolase. An enzyme implicated in crab steroidogenesis.";
RL   Endocrine 5:23-32(1996).
CC   -!- FUNCTION: ACTINS ARE HIGHLY CONSERVED PROTEINS THAT ARE INVOLVED
CC       IN VARIOUS TYPES OF CELL MOTILITY AND ARE UBIQUITOUSLY EXPRESSED
CC       IN ALL EUKARYOTIC CELLS.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS PROTEIN IS:
CC       6.8, ITS MW IS: 46 KDA.
CC   -!- SIMILARITY: BELONGS TO THE ACTIN FAMILY.
DR   INTERPRO; IPR000279; -.
DR   PROSITE; PS00406; ACTINS_1; PARTIAL.
DR   PROSITE; PS00432; ACTINS_2; PARTIAL.
DR   PROSITE; PS01132; ACTINS_ACT_LIKE; PARTIAL.
KW   Structural protein.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  976 MW;  1424005AB2CAAEB3 CRC64;

  Query Match             26.7%;  Score 12;  DB 1;  Length 8;
  Best Local Similarity   25.0%;  Pred. No. 8.8e+04;
  Matches    1;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 DINL 9
            |:::
Db        3 DVDI 6
```

```
RESULT  13
CCKN_MACEU
ID   CCKN_MACEU     STANDARD;      PRT;     8 AA.
AC   P30369;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   CHOLECYSTOKININ (CCK).
GN   CCK.
OS   Macropus eugenii (Tammar wallaby), and
OS   Dasyurus viverrinus (Southeastern quoll).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=BRAIN;
RX   MEDLINE; 88234141.
RA   Fan Z.W., Eng J., Shaw G., Yalow R.S.;
RT   "Cholecystokinin octapeptide purified from brains of Australian
RT   marsupials.";
RL   Peptides 9:429-431(1988).
CC   -!- FUNCTION: THIS PEPTIDE HORMONE INDUCES GALL BLADDER CONTRACTION
CC       AND THE RELEASE OF PANCREATIC ENZYMES IN THE GUT. ITS FUNCTION
CC       IN THE BRAIN IS NOT CLEAR.
CC   -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
DR   PIR; A43001; A43001.
DR   PIR; PQ0012; PQ0012.
DR   INTERPRO; IPR001651; -.
DR   PROSITE; PS00259; GASTRIN; 1.
KW   Amidation; Sulfatation; Hormone.
FT   MOD_RES       2      2       SULFATATION.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  1064 MW;  DDCAA68378768B5A CRC64;


  Query Match             26.7%;  Score 12;  DB 1;  Length 8;
  Best Local Similarity   33.3%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 YLSGAD 6
           |:   |
Db       2 YMGWMD 7


RESULT  14
FARD_CALVO
ID   FARD_CALVO     STANDARD;      PRT;     9 AA.
AC   P41868;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 13.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; D44787; D44787.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1028 MW;  22D10699C87AB6D8 CRC64;

  Query Match             26.7%;  Score 12;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAD 6
           :| |
Db       1 AGQD 4


RESULT  15
OXYA_SCYCA
ID   OXYA_SCYCA     STANDARD;      PRT;     9 AA.
AC   P42996;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ASVATOCIN.
OS   Scyliorhinus canicula (Spotted dogfish) (Spotted catshark).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Galeomorphii; Galeoidea; Carcharhiniformes;
OC   Scyliorhinidae; Scyliorhinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PITUITARY;
RX   MEDLINE; 95062247.
RA   Chauvet J., Rouille Y., Chauveau C., Chauvet M.-T., Acher R.;
RT   "Special evolution of neurohypophysial hormones in cartilaginous
RT   fishes: asvatocin and phasvatocin, two oxytocin-like peptides
RT   isolated from the spotted dogfish (Scyliorhinus caniculus).";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:11266-11270(1994).
CC   -!- FUNCTION: DISPLAYS OXYTOCIC ACTIVITY ON RAT UTERUS.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  982 MW;  17EDD76EB444404B CRC64;


  Query Match             26.7%;  Score 12;  DB 1;  Length 9;
  Best Local Similarity   14.3%;  Pred. No. 8.8e+04;
  Matches    1;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 YLSGADI 7
           |::    :
Db       2 YINNCPV 8


RESULT  16
OXYT_BUFRE
ID   OXYT_BUFRE     STANDARD;      PRT;     9 AA.
AC   P42995;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   SERITOCIN.
OS   Bufo regularis (African toad).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Bufonidae;
OC   Bufo.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PITUITARY NEUROINTERMEDIATE LOBE;
RX   MEDLINE; 96059313.
RA   Chauvet J., Michel G., Ouedraogo Y., Chou J., Chait B.T., Acher R.;
RT   "A new neurohypophysial peptide, seritocin ([Ser5,Ile8]-oxytocin),
RT   identified in a dryness-resistant African toad, Bufo regularis.";
RL   Int. J. Pept. Protein Res. 45:482-487(1995).
CC   -!- FUNCTION: DEVOID OF OXYTOCIC ACTIVITY.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
```

```
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; FALSE_NEG.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  983 MW;  17FF476EA5A6D04B CRC64;

  Query Match             26.7%;  Score 12;  DB 1;  Length 9;
  Best Local Similarity   28.6%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 YLSGADI 7
          |:    |
Db      2 YIQSCPI 8


RESULT  17
PGLR_DIAAB
ID   PGLR_DIAAB     STANDARD;      PRT;     9 AA.
AC   P81179;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ENDO-POLYGALACTURONASE (PG) (EC 3.2.1.15) (FRAGMENT).
OS   Diaprepes abbreviatus (Sugarcane rootstalk borer weevil).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC   Cucujiformia; Curculionidae; Entiminae; Entimini; Diaprepes.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LARVAL GUT;
RA   Doostdar H., McCollum T.G., Mayer R.T.;
RT   "Purification and characterization of an endo-polygalacturonase from
RT   the gut of West Indies sugarcane rootstalk borer weevil (Diaprepes
RT   abbreviatus L.) larvae.";
RL   Comp. Biochem. Physiol. 118B:861-867(1997).
CC   -!- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-ALPHA-D-
CC       GALACTOSIDURONIC LINKAGES IN PECTATE AND OTHER GALACTURONANS.
CC   -!- INDUCTION: INHIBITED BY CITRUS PGIP.
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS PROTEIN IS:
CC       9.4, ITS MW IS: 44.5 KDA.
CC   -!- SIMILARITY: WEAK TO OTHER POLYGALACTURONASES.
KW   Hydrolase; Glycosidase; Cell wall.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1041 MW;  1F49087042DB41BB CRC64;

  Query Match             26.7%;  Score 12;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSG 4
          |: |
Db      4 YVIG 7


RESULT  18
UPA6_HUMAN
ID   UPA6_HUMAN     STANDARD;      PRT;     9 AA.
AC   P30092;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE OF PLASMA (SPOT 14) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PLASMA;
RX   MEDLINE; 93092937.
```

```
RA   Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,
RA   Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,
RA   Hochstrasser D.F.;
RT   "Plasma protein map: an update by microsequencing.";
RL   Electrophoresis 13:707-714(1992).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 5, ITS MW IS: 48 KDA.
DR   SWISS-2DPAGE; P30092; HUMAN.
FT   NON_TER       1      1
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  935 MW;  5282F2CAA8676447 CRC64;

  Query Match             26.7%;  Score 12;  DB 1;  Length 9;
  Best Local Similarity   33.3%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 LSGADI 7
          |:  |:
Db      2 LNPGDV 7


RESULT  19
FAR3_HIRME
ID   FAR3_HIRME     STANDARD;      PRT;     4 AA.
AC   P42562;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE YLRF-AMIDE.
OS   Hirudo medicinalis (Medicinal leech).
OC   Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;
OC   Arynchobdellida; Hirudiniformes; Hirudinidae; Hirudo.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92195954.
RA   Evans B.D., Pohl J., Kartsonis M.A., Calabrese R.L.;
RT   "Identification of RFamide neuropeptides in the medicinal leech.";
RL   Peptides 12:897-908(1991).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       4      4       AMIDATION.
SQ   SEQUENCE   4 AA;  598 MW;  69D4073B30000000 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YL 2
          ||
Db      1 YL 2


RESULT  20
PRCT_PERAM
ID   PRCT_PERAM     STANDARD;      PRT;     5 AA.
AC   P01373;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   PROCTOLIN.
OS   Periplaneta americana (American cockroach),
OS   Limulus polyphemus (Atlantic horseshoe crab), and
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blattoidea; Blattidae; Periplaneta.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=P.AMERICANA;
RX   MEDLINE; 76074708.
```

```
RA   Starratt A.N., Brown B.E.;
RT   "Structure of the pentapeptide proctolin, a proposed neurotransmitter
RT   in insects.";
RL   Life Sci. 17:1253-1256(1975).
RN   [2]
RP   BIOLOGICAL SOURCE.
RC   SPECIES=P.AMERICANA;
RX   MEDLINE; 81225865.
RA   O'Shea M., Adams M.E.;
RT   "Pentapeptide (proctolin) associated with an identified neuron.";
RL   Science 213:567-569(1981).
RN   [3]
RP   SEQUENCE.
RC   SPECIES=L.POLYPHEMUS;
RX   MEDLINE; 90287800.
RA   Groome J.R., Tillinghast E.K., Townley M.A., Vetrovs A.,
RA   Watson W.H. III, Hunt D.F., Griffin P.R., Alexander J.E.,
RA   Shabanowitz J.;
RT   "Identification of proctolin in the central nervous system of the
RT   horseshoe crab, Limulus polyphemus.";
RL   Peptides 11:205-211(1990).
RN   [4]
RP   SEQUENCE.
RC   SPECIES=C.MAENAS;
RX   MEDLINE; 86232789.
RA   Stangier J., Dircksen H., Keller R.;
RT   "Identification and immunocytochemical localization of proctolin in
RT   pericardial organs of the shore crab, Carcinus maenas.";
RL   Peptides 7:67-72(1986).
CC   -!- FUNCTION: STIMULATES CARDIAC OUTPUT AND HINDGUT MOTILITY,
CC       MODULATES VISCERAL AND SKELETAL MUSCLE IN MANY ARTHROPODS.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE LATERAL WHITE NEURONS AND IN
CC       THE CRAB PERICARDIAL ORGANS.
DR   PIR; A01644; HOROHA.
DR   PIR; A60411; A60411.
KW   Neuropeptide.
SQ   SEQUENCE   5 AA;  649 MW;  71B7673B44600000 CRC64;


  Query Match             24.4%;  Score 11;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YL 2
            ||
Db        2 YL 3


RESULT  21
FAR2_ASCSU
ID   FAR2_ASCSU     STANDARD;      PRT;     7 AA.
AC   P31890;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE AF2.
OS   Ascaris suum (Pig roundworm) (Ascaris lumbricoides), and
OS   Panagrellus redivivus.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC   Ascarididae; Ascaris.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=A.SUUM;
RX   MEDLINE; 93324431.
RA   Cowden C., Stretton A.O.W.;
RT   "AF2, an Ascaris neuropeptide: isolation, sequence, and bioactivity.";
RL   Peptides 14:423-430(1993).
RN   [2]
RP   SEQUENCE.
RC   SPECIES=P.REDIVIVUS;
RX   MEDLINE; 95060998.
RA   Maule A.G., Shaw C., Bowman J.W.;
RT   "The FMRFamide-like neuropeptide AF2 (Ascaris suum) is present in the
```

```
RT   free-living nematode, Panagrellus redivivus (Nematoda, Rhabditida).";
RL   Parasitology 109:351-356(1994).
CC   -!- FUNCTION: HAS EFFECTS ON MUSCLE TENSION.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE NERVE CORDS AND A VARIETY OF
CC       GANGLIA PARTICULARLY IN THE ANTERIOR REGIONS.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  992 MW;  69D4073B5B11E350 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YL 2
            ||
Db        4 YL 5


RESULT  22
GFRP_MOUSE
ID   GFRP_MOUSE     STANDARD;      PRT;     7 AA.
AC   P99025;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GTP CYCLOHYDROLASE I FEEDBACK REGULATORY PROTEIN (P35) (FRAGMENT).
GN   GFRP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LIVER;
RA   Sanchez J.-C., Rouge V., Frutiger S., Hughes G., Yan J.X.,
RA   Hoogland C., Appel R.D., Binz P.-A., Hochstrasser D.F.,
RA   Cowthorne M.;
RL   Submitted (AUG-1998) to the SWISS-PROT data bank.
CC   -!- FUNCTION: MEDIATES TETRAHYDROBIOPTERIN INHIBITION OF GTP
CC       CYCLOHYDROLASE I. THIS INHIBITION IS REVERSED BY L-PHENYLALANINE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
DR   SWISS-2DPAGE; P99025; MOUSE.
FT   INIT_MET      0      0
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  806 MW;  71B5B057273B4700 CRC64;


  Query Match             24.4%;  Score 11;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YL 2
            ||
Db        2 YL 3


RESULT  23
CAD1_ENTFA
ID   CAD1_ENTFA     STANDARD;      PRT;     8 AA.
AC   P13268;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   SEX PHEROMONE CAD1.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 85051889.
```

```
RA   Mori M., Sagakami Y., Narita M., Isogai A., Fujino M., Kitada C.,
RA   Craig R.A., Clewell D.B., Suzuki A.;
RT   "Isolation and structure of the bacterial sex pheromone, cAD1, that
RT   induces plasmid transfer in Streptococcus faecalis.";
RL   FEBS Lett. 178:97-100(1984).
CC   -!- FUNCTION: CAD1 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE
CC       HEMOLYSIN PLASMID PAD1.
KW   Pheromone.
SQ   SEQUENCE   8 AA;  819 MW;  047DD732C735B9C7 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LSG 4
              |:|
Db          6 LAG 8


RESULT  24
FAR8_CALVO
ID   FAR8_CALVO     STANDARD;      PRT;     8 AA.
AC   P41863;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 8.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; H41978; H41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  957 MW;  72D40699CAA44DD8 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 GAD 6
              ||:
Db          1 GAN 3


RESULT  25
D1_NEPNO
ID   D1_NEPNO       STANDARD;      PRT;     9 AA.
AC   P24816;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   01-MAR-1992 (Rel. 21, Last annotation update)
DE   GASTRIN/CHOLECYSTOKININ-LIKE PEPTIDE D1.
OS   Nephrops norvegicus (Norway lobster).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC   Nephropoidea; Nephropidae; Nephrops.
RN   [1]
RP   SEQUENCE.
```

```
RC   TISSUE=STOMACH;
RX   MEDLINE; 92082847.
RA   Favrel P., Kegel G., Sedlmeier D., Keller R., van Wormhoudt A.;
RT   "Structure and biological activity of crustacean gastrointestinal
RT   peptides identified with antibodies to gastrin/cholecystokinin.";
RL   Biochimie 73:1233-1239(1991).
CC   -!- FUNCTION: MAY CONTROL DIGESTION PROCESSES IN CRUSTACEANS.
CC   -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
DR   PIR; A48398; A48398.
KW   Hormone.
SQ   SEQUENCE   9 AA;  1038 MW;  60EC79CAB6D8787B CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 GAD 6
              | |
Db          4 GQD 6


RESULT  26
FAR5_CALVO
ID   FAR5_CALVO     STANDARD;      PRT;     9 AA.
AC   P41860;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 5.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; E41978; E41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1068 MW;  39D10699CAB6D867 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 GAD 6
              | |
Db          3 GQD 5


RESULT  27
FAR7_CALVO
ID   FAR7_CALVO     STANDARD;      PRT;     9 AA.
AC   P41862;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 7.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
```

```
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; G41978; G41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1081 MW;  E1D10699CAB6D86A CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GAD 6
           | |
Db       3 GQD 5


RESULT  28
MOSF_CLYJA
ID   MOSF_CLYJA     STANDARD;      PRT;     9 AA.
AC   P19853;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   [PHE-6]-MOSACT.
OS   Clypeaster japonicus (Sand dollar).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Gnathostomata; Clypeasteroida;
OC   Clypeasteridae; Clypeaster.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=EGG JELLY;
RA   Suzuki N., Kurita M., Yoshino K.I., Kajiura H., Nomura K.,
RA   Yamaguchi M.;
RT   "Purification and structure of mosact and its derivatives from the
RT   egg jelly of the sea urchin Clypeaster japonicus.";
RL   Zool. Sci. 4:649-656(1987).
CC   -!- FUNCTION: STIMULATES SPERM RESPIRATION AND MOTILITY.
DR   PIR; JN0027; JN0027.
SQ   SEQUENCE   9 AA;  924 MW;  93245729CDC5BAB5 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           :| |
Db       6 FLIG 9


RESULT  29
OXYA_SQUAC
ID   OXYA_SQUAC     STANDARD;      PRT;     9 AA.
AC   P42999;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ASPARTOCIN (ASPARGTOCIN).
OS   Squalus acanthias (Spiny dogfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Squaloidei; Squalidae; Squalus.
RN   [1]
```

```
RP   SEQUENCE.
RX   MEDLINE; 73031727.
RA   Acher R., Chauvet J., Chauvet M.-T.;
RT   "Phylogeny of the neurohypophysial hormones. Two new active peptides
RT   isolated from a cartilaginous fish, Squalus acanthias.";
RL   Eur. J. Biochem. 29:12-19(1972).
RN   [2]
RP   SEQUENCE.
RX   MEDLINE; 72128038.
RA   Acher R., Chauvet J., Chauvet M.-T., Fontaine M.;
RT   "Identification of 2 new neurohypophyseal hormones, valitocin (Val8-
RT   oxytocin) and aspartocin (Asn4-oxytocin) in a selachian fish, the
RT   spiny dog-fish (Squalus acanthias).";
RL   C. R. Acad. Sci., D, Sci. Nat. 274:313-316(1972).
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  996 MW;  17F8376EB444404B CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   14.3%;  Pred. No. 8.8e+04;
  Matches    1;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 YLSGADI 7
           |::   :
Db       2 YINNCPL 8


RESULT  30
OXYV_SQUAC
ID   OXYV_SQUAC     STANDARD;      PRT;     9 AA.
AC   P43000;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   VALITOCIN.
OS   Squalus acanthias (Spiny dogfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Squaloidei; Squalidae; Squalus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 73031727.
RA   Acher R., Chauvet J., Chauvet M.-T.;
RT   "Phylogeny of the neurohypophysial hormones. Two new active peptides
RT   isolated from a cartilaginous fish, Squalus acanthias.";
RL   Eur. J. Biochem. 29:12-19(1972).
RN   [2]
RP   SEQUENCE.
RX   MEDLINE; 72128038.
RA   Acher R., Chauvet J., Chauvet M.-T., Fontaine M.;
RT   "Identification of 2 new neurohypophyseal hormones, valitocin (Val8-
RT   oxytocin) and aspartocin (Asn4-oxytocin) in a selachian fish, the
RT   spiny dog-fish (Squalus acanthias).";
RL   C. R. Acad. Sci., D, Sci. Nat. 274:313-316(1972).
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  996 MW;  17EDD76EB456D04B CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   14.3%;  Pred. No. 8.8e+04;
  Matches    1;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy        1 YLSGADI 7
            |:    :
Db        2 YIQNCPV 8


RESULT  31
ACH1_ACHFU
ID   ACH1_ACHFU     STANDARD;      PRT;     4 AA.
AC   P35904;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   ACHATIN-I.
OS   Achatina fulica (Giant African snail).
OC   Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC   Achatinacea; Achatinidae; Achatina.
RN   [1]
RP   SEQUENCE, CHARACTERIZATION, AND SYNTHESIS.
RC   STRAIN=FERUSSAC; TISSUE=GANGLION;
RX   MEDLINE; 89273551.
RA   Kamatani Y., Minakata H., Kenny P.T.M., Iwashita T., Watanabe K.,
RA   Funase K., Sun X.P., Yongsiri A., Kim K.H., Novales-Li P.,
RA   Novales E.T., Kanapi C.G., Takeuchi H., Nomoto K.;
RT   "Achatin-I, an endogenous neuroexcitatory tetrapeptide from Achatina
RT   fulica Ferussac containing a D-amino acid residue.";
RL   Biochem. Biophys. Res. Commun. 160:1015-1020(1989).
RN   [2]
RP   CHARACTERIZATION.
RC   STRAIN=FERUSSAC; TISSUE=HEART ATRIUM;
RX   MEDLINE; 91264856.
RA   Fujimoto K., Kubota I., Yasuda-Kamatani Y., Minakata H., Nomoto K.,
RA   Yoshida M., Harada A., Muneoka Y., Kobayashi M.;
RT   "Purification of achatin-I from the atria of the African giant snail,
RT   Achatina fulica, and its possible function.";
RL   Biochem. Biophys. Res. Commun. 177:847-853(1991).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY.
RX   MEDLINE; 93014529.
RA   Ishida T., In Y., Doi M., Inoue M., Yasuda-Kamatani Y., Minakata H.,
RA   Iwashita T., Nomoto K.;
RT   "Crystal structure and molecular conformation of achatin-I
RT   (H-Gly-D-Phe-Ala-Asp-OH), an endogenous neuropeptide containing a
RT   D-amino acid residue.";
RL   Int. J. Pept. Protein Res. 39:258-264(1992).
CC   -!- FUNCTION: NEUROEXCITATORY PEPTIDE; INCREASES THE IMPULSE FREQUENCY
CC       AND PRODUCES A SPIKE BROADENING OF THE IDENTIFIED HEART EXCITATORY
CC       NEURON (PON); ALSO ENHANCES THE AMPLITUDE AND FREQUENCY OF THE
CC       HEART BEAT. HAS ALSO AN EFFECT ON SEVERAL OTHER MUSCLES.
DR   PIR; A32480; A32480.
KW   Hormone; D-amino acid.
FT   MOD_RES       2      2       D-PHENYLALANINE.
SQ   SEQUENCE   4 AA;  408 MW;  6AADD9C810000000 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 AD 6
            ||
Db        3 AD 4


RESULT  32
UXA4_CHLTR
ID   UXA4_CHLTR     STANDARD;      PRT;     5 AA.
AC   P38005;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE FROM ELEMENTARY BODY (FRAGMENT).
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE.
RC   STRAIN=L2/434/BU;
RA   Bini L., Santucci A., Magi B., Marzocchi B., Sanchez-Campillo M.,
RA   Comanducci M., Christianen G., Birkelund S., Vtretou E., Ratti G.,
RA   Pallini V.;
RL   Submitted (SEP-1994) to the SWISS-PROT data bank.
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 4.5, ITS MW IS: 28 KDA.
FT   NON_TER       5      5
SQ   SEQUENCE   5 AA;  474 MW;  75BAA865AA800000 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 SG 4
            ||
Db        2 SG 3


RESULT  33
CIP2_MYTED
ID   CIP2_MYTED     STANDARD;      PRT;     6 AA.
AC   P13737;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   CONTRACTION-INHIBITING PEPTIDE II (MIP II).
OS   Mytilus edulis (Blue mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytiloidea; Mytilidae; Mytilus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PEDAL GANGLION;
RX   MEDLINE; 88240357.
RA   Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.;
RT   "Structures and actions of Mytilus inhibitory peptides.";
RL   Biochem. Biophys. Res. Commun. 152:1376-1382(1988).
CC   -!- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN
CC       MUSCLES.
CC   -!- SIMILARITY: TO MIP I.
DR   PIR; B27696; B27696.
KW   Hormone; Amidation.
FT   MOD_RES       6      6       AMIDATION.
SQ   SEQUENCE   6 AA;  621 MW;  72C9C6876DD81000 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        4 GA 5
            ||
Db        1 GA 2


RESULT  34
TMOF_SARBU
ID   TMOF_SARBU     STANDARD;      PRT;     6 AA.
AC   P41495;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   TRYPSIN-MODULATING OOSTATIC FACTOR (TMOF).
OS   Sarcophaga bullata (Grey flesh fly) (Neobellieria bullata).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Sarcophagidae; Sarcophaga.
RN   [1]
RP   SEQUENCE, AND SYNTHESIS.
RC   TISSUE=OVARY;
```

```
RX   MEDLINE; 94211930.
RA   Bylemans D., Borovsky D., Hunt D.F., Shabanowitz J., Grauwels L.,
RA   de Loof A.;
RT   "Sequencing and characterization of trypsin modulating oostatic
RT   factor (TMOF) from the ovaries of the grey fleshfly, Neobellieria
RT   (Sarcophaga) bullata.";
RL   Regul. Pept. 50:61-72(1994).
CC   -!- FUNCTION: HAS AN OOSTATIC ACTIVITY. INHIBITS TRYPSIN BIOSYNTHESIS
CC       IN THE MIDGUT WHICH INDIRECTLY REDUCES THE VITELLOGENIN
CC       CONCENTRATION IN THE HEMOLYMPH RESULTING IN INHIBITION OF OOCYTE
CC       DEVELOPMENT.
CC   -!- DEVELOPMENTAL STAGE: SYNTHESIZED AND RELEASED FROM FOLLICULAR
CC       EPITHELIUM AFTER A BLOOD MEAL.
KW   Hormone.
SQ   SEQUENCE   6 AA;  695 MW;  61E72451B7642000 CRC64;


  Query Match             22.2%;  Score 10;  DB 1;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        8 NL 9
            ||
Db        4 NL 5


RESULT  35
UN06_CLOPA
ID   UN06_CLOPA     STANDARD;      PRT;     6 AA.
AC   P81351;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   UNKNOWN PROTEIN CP 6 FROM 2D-PAGE (FRAGMENT).
OS   Clostridium pasteurianum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
RN   [1]
RP   SEQUENCE.
RC   STRAIN=W5;
RX   MEDLINE; 98291870.
RA   Flengsrud R., Skjeldal L.;
RT   "Two-dimensional gel electrophoresis separation and N-terminal
RT   sequence analysis of proteins from Clostridium pasteurianum W5.";
RL   Electrophoresis 19:802-806(1998).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 5.0, ITS MW IS: 75.9 KDA.
FT   NON_TER       6      6
SQ   SEQUENCE   6 AA;  657 MW;  605B1DC1A45A8000 CRC64;


  Query Match             22.2%;  Score 10;  DB 1;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 ADI 7
            |:|
Db        4 AEI 6


Search completed: December 16, 2000, 04:23:31
Job time: 4566 sec
```

```
                       GenCore version 4.5
             Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 03:35:12 ; Search time 89.11 Seconds
                                             (without alignments)
                                             6.409 Million cell updates/sec

Title:          US-09-529-121-3
Perfect score:  45
Sequence:       1 YLSGADINL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters:      793

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       PIR_65:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                %
Result        Query
  No.   Score Match Length DB  ID                   Description
--------------------------------------------------------------------------
    1      16  35.6      7  2  PT0542               T-cell receptor be
    2      16  35.6      8  2  A21440               variant surface gl
    3      16  35.6      8  2  JS0318               leucokinin VIII -
    4      16  35.6      9  2  A61364               isotocin - common
    5      16  35.6      9  4  I57650               hemoglobin alpha c
    6      15  33.3      5  2  PT0540               T-cell receptor be
    7      15  33.3      6  2  PT0726               T-cell receptor be
    8      15  33.3      7  2  PT0526               T-cell receptor be
    9      15  33.3      7  2  PT0676               T-cell receptor be
   10      15  33.3      8  2  T13818               cytochrome oxidase
   11      15  33.3      9  2  F41978               calliFMRFamide 6 -
   12      15  33.3      9  2  PT0288               Ig heavy chain CRD
   13      15  33.3      9  2  G41946               T-cell receptor ga
   14      14  31.1      5  2  PT0679               T-cell receptor be
   15      14  31.1      6  2  PT0605               T-cell receptor be
   16      14  31.1      6  2  PT0593               T-cell receptor be
   17      14  31.1      7  2  PC2370               probable H+-transp
   18      14  31.1      7  2  PT0654               T-cell receptor be
   19      14  31.1      7  2  PT0722               T-cell receptor be
   20      14  31.1      8  2  S65647               2-hydroxyglutaryl-
   21      14  31.1      8  2  PN0043               phosphatidylethano
   22      14  31.1      8  2  PT0557               T-cell receptor be
   23      14  31.1      9  1  AKLQIM               locustamyoinhibiti
   24      14  31.1      9  2  A57444               neuropeptide Grb-A
   25      14  31.1      9  2  B57444               neuropeptide Grb-A
   26      14  31.1      9  2  C57444               neuropeptide Grb-A
   27      14  31.1      9  2  PT0268               Ig heavy chain CRD
   28      13  28.9      4  2  S43959               Ig mu chain V regi
   29      13  28.9      7  2  I50210               gene c-rel protein
   30      13  28.9      7  2  A58718               carnocin UI49 - Ca
   31      13  28.9      8  2  A41117               acetylcholinestera
   32      13  28.9      8  2  PT0547               T-cell receptor be
   33      13  28.9      9  2  QDRB                 delta sleep-induci
   34      13  28.9      9  2  C41170               photosystem II pro
   35      13  28.9      9  2  A61386               macrophage inhibit
   36      13  28.9      9  2  PH0935               T-cell receptor be
   37      13  28.9      9  2  PH0918               T-cell receptor be
   38      12  26.7      4  2  A26209               protein-glutamine
   39      12  26.7      5  2  S62883               seminal plasma pro
   40      12  26.7      6  2  I51434               H4 histone - Afric
   41      12  26.7      7  2  S16364               opacity protein P.
   42      12  26.7      7  2  B35890               RNA-directed DNA p
   43      12  26.7      7  2  S20446               elastase - Pseudom
   44      12  26.7      7  2  S29735               polyphosphate--glu
   45      12  26.7      7  2  A34818               vicilin 72K chain
   46      12  26.7      8  2  PQ0012               cholecystokinin -
   47      12  26.7      8  2  A43001               cholecystokinin -
   48      12  26.7      8  2  PQ0701               unidentified 6.5/3
   49      12  26.7      8  2  PL0184               capsid protein VP-
   50      12  26.7      9  2  D44787               calliFMRFamide 13
   51      11  24.4      5  1  HOROHA               proctolin - Americ
   52      11  24.4      5  2  A41225               copper resistance
   53      11  24.4      5  2  B31836               20K protein - Rick
   54      11  24.4      5  2  A60411               proctolin - Atlant
   55      11  24.4      5  2  PT0267               Ig heavy chain CRD
   56      11  24.4      5  2  C23751               spinal cord peptid
   57      11  24.4      5  2  PT0651               T-cell receptor be
   58      11  24.4      6  2  B44510               hypothetical prote
   59      11  24.4      6  2  PT0280               Ig heavy chain CRD
   60      11  24.4      7  2  S09066               globulin IV alpha
   61      11  24.4      7  2  S78024               ribosomal protein
   62      11  24.4      7  2  PT0620               T-cell receptor be
   63      11  24.4      8  2  A05169               neuropeptide M-I -
   64      11  24.4      8  2  H41978               calliFMRFamide 8 -
   65      11  24.4      8  2  E60588               sperm-activating p
   66      11  24.4      8  2  A61467               penalbumin - Adeli
   67      11  24.4      8  2  PT0311               Ig heavy chain CRD
   68      11  24.4      8  2  PT0627               T-cell receptor be
   69      11  24.4      8  2  PT0522               T-cell receptor be
   70      11  24.4      8  2  B47594               aspartate kinase (
   71      11  24.4      8  2  S68325               blood cell protein
   72      11  24.4      9  2  A44873               caldesmon - rabbit
   73      11  24.4      9  2  A43848               cell surface adhes
   74      11  24.4      9  2  E41978               calliFMRFamide 5 -
   75      11  24.4      9  2  G41978               calliFMRFamide 7 -


                              ALIGNMENTS


RESULT   1
PT0542
T-cell receptor beta chain V-D-J region (126-1BA) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0542
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0542
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-7 <FEE>
A;Experimental source: day 18 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             35.6%;  Score 16;  DB 2;  Length 7;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy          3 SGAD 6
              || |
Db          2 SGGD 5


RESULT    2
A21440
variant surface glycoprotein pSLc1 - Trypanosoma brucei (fragment)
C;Species: Trypanosoma brucei
C;Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 20-Mar-1998
C;Accession: A21440
R;Parsons, M.; Nelson, R.G.; Watkins, K.P.; Agabian, N.
Cell 38, 309-316, 1984
A;Title: Trypanosome mRNAs share a common 5' spliced leader sequence.
A;Reference number: A90853; MUID:84282716
A;Accession: A21440
A;Molecule type: mRNA
A;Residues: 1-8 <PAR>
A;Cross-references: GB:K02195; NID:g162150; PID:g162151
C;Keywords: glycoprotein


  Query Match             35.6%;  Score 16;  DB 2;  Length 8;
  Best Local Similarity   33.3%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LSGADI 7
              :|| ::
Db          1 MSGKEV 6


RESULT    3
JS0318
leucokinin VIII - Madeira cockroach
C;Species: Leucophaea maderae (Madeira cockroach)
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 20-Jun-2000
C;Accession: JS0318
R;Holman, G.M.; Cook, B.J.; Nachman, R.J.
Comp. Biochem. Physiol. C 88, 31-34, 1987
A;Title: Isolation, primary structure and synthesis of leucokinins VII and VIII: the fir
A;Reference number: JS0317
A;Accession: JS0318
A;Molecule type: protein
A;Residues: 1-8 <HOL>
C;Comment: Leucokinins, a family of cephalomyotropic peptides, stimulate contractile act
C;Keywords: amidated carboxyl end; cephalomyotropic peptide
F;8/Modified site: amidated carboxyl end (Gly) #status experimental


  Query Match             35.6%;  Score 16;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 GAD 6
              |||
Db          1 GAD 3


RESULT    4
A61364
isotocin - common carp
C;Species: Cyprinus carpio (common carp)
C;Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 20-Jun-2000
C;Accession: A61364
R;Acher, R.; Chauvet, J.; Chauvet, M.T.; Crepy, D.
Comp. Biochem. Physiol. A 14, 245-254, 1965
A;Title: Caracterisation des hormones neurohypophysaires d'un poisson osseux d'eau douce
A;Reference number: A61364
A;Accession: A61364
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <ACH>
C;Superfamily: oxytocin-neurophysin
C;Keywords: amidated carboxyl end; neuropeptide; posterior pituitary
F;9/Modified site: amidated carboxyl end (Gly) #status experimental


  Query Match             35.6%;  Score 16;  DB 2;  Length 9;
  Best Local Similarity   42.9%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 YLSGADI 7
              |:|   |
Db          2 YISNCPI 8


RESULT    5
I57650
hemoglobin alpha chain - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 31-Jul-1997 #text_change 20-Apr-2000
C;Accession: I57650
R;Whitelaw, E.; Hogben, P.; Hanscombe, O.; Proudfoot, N.J.
Mol. Cell. Biol. 9, 241-251, 1989
A;Title: Transcriptional promiscuity of the human alpha-globin gene.
A;Reference number: I57650; MUID:89181576
A;Accession: I57650
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-9 <WHI>
A;Cross-references: GB:M23454; NID:g340922; PIDN:AAA52629.1; PID:g553329
A;Note: engineered sequence; this sequence was not determined in this report
C;Genetics:
A;Gene: GDB:HBA1
A;Cross-references: GDB:119293
A;Map position: 16p13.3-16p13.3


  Query Match             35.6%;  Score 16;  DB 4;  Length 9;
  Best Local Similarity   80.0%;  Pred. No. 1.8e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LSGAD 6
              || ||
Db          3 LSPAD 7


RESULT    6
PT0540
T-cell receptor beta chain V-D-J region (126-1L) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0540
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0540
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-5 <FEE>
A;Experimental source: day 18 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             33.3%;  Score 15;  DB 2;  Length 5;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 SGAD 6
              || |
Db          2 SGED 5


RESULT    7
PT0726
```

```
T-cell receptor beta chain V-D-J region (161-2D) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0726
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0726
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-6 <FEE>
A;Experimental source: newborn thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             33.3%;  Score 15;  DB 2;  Length 6;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAD 6
           || |
Db       2 SGED 5

RESULT   8
PT0526
T-cell receptor beta chain V-D-J region (100-4E) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0526
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0526
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-7 <FEE>
A;Experimental source: adult thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             33.3%;  Score 15;  DB 2;  Length 7;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAD 6
           || |
Db       2 SGED 5

RESULT   9
PT0676
T-cell receptor beta chain V-D-J region (140-1AL) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0676
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0676
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-7 <FEE>
A;Experimental source: day 18 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             33.3%;  Score 15;  DB 2;  Length 7;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy       3 SGAD 6
           || |
Db       2 SGED 5

RESULT  10
T13818
cytochrome oxidase subunit I - Atlantic hagfish mitochondrion (fragment)
C;Species: mitochondrion Myxine glutinosa (Atlantic hagfish)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C;Accession: T13818
R;Delarbre, C.; Barriel, V.; Tillier, S.; Janvier, P.; Gachelin, G.
Mol. Biol. Evol. 14, 807-813, 1997
A;Title: The main features of the craniate mitochondrial DNA between the ND1 and the
A;Reference number: Z17775; MUID:97398704
A;Accession: T13818
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-8 <DEL>
A;Cross-references: EMBL:Y09527; NID:e1011465; PID:e329906; PIDN:CAA70718.1
C;Genetics:
A;Genome: mitochondrion
A;Note: COI
C;Keywords: mitochondrion


  Query Match             33.3%;  Score 15;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLS 3
           |||
Db       2 YLS 4

RESULT  11
F41978
calliFMRFamide 6 - bluebottle fly (Calliphora vomitoria)
C;Species: Calliphora vomitoria
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C;Accession: F41978
R;Duve, H.; Johnsen, A.H.; Sewell, J.C.; Scott, A.G.; Orchard, I.; Rehfeld, J.F.; Tho
Proc. Natl. Acad. Sci. U.S.A. 89, 2326-2330, 1992
A;Title: Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH-2 neuropeptides (d
A;Reference number: A41978; MUID:92196111
A;Accession: F41978
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <DUV>
C;Keywords: amidated carboxyl end; neuropeptide
F;9/Modified site: amidated carboxyl end (Phe) #status experimental


  Query Match             33.3%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAD 6
           || |
Db       2 SGQD 5

RESULT  12
PT0288
Ig heavy chain CRD3 region (clone 4-106) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C;Accession: PT0288
R;Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991
A;Title: Preferential utilization of specific immunoglobulin heavy chain diversity an
A;Reference number: PT0222; MUID:91108337
```

```
A;Accession: PT0288
A;Molecule type: DNA
A;Residues: 1-9 <YAM>
A;Experimental source: B lymphocyte
C;Keywords: heterotetramer; immunoglobulin


  Query Match             33.3%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSG 4
              | ||
Db          5 YSSG 8

RESULT  13
G41946
T-cell receptor gamma chain (2t.23) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 07-May-1999
C;Accession: G41946
R;Whetsell, M.; Mosley, R.L.; Whetsell, L.; Schaefer, F.V.; Miller, K.S.; Klein, J.R.
Mol. Cell. Biol. 11, 5902-5909, 1991
A;Title: Rearrangement and junctional-site sequence analyses of T-cell receptor gamma ge
A;Reference number: A41946; MUID:92049316
A;Accession: G41946
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-9 <WHE>
C;Keywords: T-cell receptor


  Query Match             33.3%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSG 4
              | ||
Db          5 YSSG 8

RESULT  14
PT0679
T-cell receptor beta chain V-D-J region - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0679; PT0708
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0679
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-5 <FEE>
A;Experimental source: day 18 fetal thymus, strain BALB/c, 154-2J
A;Accession: PT0708
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-5 <FE2>
A;Experimental source: newborn thymus, strain BALB/c, 161-2B
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 5;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 SGAD 6
              || |
Db          2 SGDD 5
```

```
RESULT  15
PT0605
T-cell receptor beta chain V-D-J region (120-1L) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0605
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0605
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-6 <FEE>
A;Experimental source: newborn thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 SGA 5
              |||
Db          2 SGA 4

RESULT  16
PT0593
T-cell receptor beta chain V-D-J region (159-1F) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0593
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0593
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-6 <FEE>
A;Experimental source: day 19 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 SGA 5
              |||
Db          4 SGA 6

RESULT  17
PC2370
probable H+-transporting ATP synthase (EC 3.6.1.34) alpha chain [similarity] - Bacill
N;Alternate names: unidentified 78K protein
C;Species: Bacillus cereus
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: PC2370
R;Matsuno, K.; Miyamoto, T.; Yamaguchi, K.; Sayed, M.A.; Kajiwara, T.; Hatano, S.
Biosci. Biotechnol. Biochem. 59, 231-235, 1995
A;Title: Identification of DNA-binding proteins changed after induction of sporulatio
A;Reference number: PC2369; MUID:95218265
A;Accession: PC2370
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-7 <MAS>
C;Keywords: ATP biosynthesis; hydrolase
```

```
  Query Match             31.1%;  Score 14;  DB 2;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DIN 8
           |:|
Db       2 DLN 4

RESULT  18
PT0654
T-cell receptor beta chain V-D-J region (121-1BK) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0654
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0654
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-7 <FEE>
A;Experimental source: day 4 postnatal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGA 5
           |||
Db       2 SGA 4

RESULT  19
PT0722
T-cell receptor beta chain V-D-J region (135-1G) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0722
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0722
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-7 <FEE>
A;Experimental source: newborn thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 7;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAD 6
           || |
Db       2 SGDD 5

RESULT  20
S65647
2-hydroxyglutaryl-CoA dehydratase - Acidaminococcus fermentans (fragment)
C;Species: Acidaminococcus fermentans
C;Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 07-May-1999
C;Accession: S65647
R;Mueller, U.; Buckel, W.
Eur. J. Biochem. 230, 698-704, 1995
A;Title: Activation of (R)-2-hydroxyglutaryl-CoA dehydratase from Acidaminococcus fermen
A;Reference number: S65647; MUID:95331308
```

```
A;Accession: S65647
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-8 <MUE>

  Query Match             31.1%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GADI 7
           | |:
Db       4 GIDV 7

RESULT  21
PN0043
phosphatidylethanol amine-binding protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 29-Oct-1997 #sequence_revision 29-Oct-1997 #text_change 23-Jan-1998
C;Accession: PN0043
R;Kato, H.
Kawasaki Igakkaishi 22, 245-259, 1996
A;Title: Analysis of proteins isolated by two dimensional electrophoresis of mouse ne
A;Reference number: PN0041
A;Accession: PN0043
A;Molecule type: protein
A;Residues: 1-8 <KAT>
A;Experimental source: neuroblastoma cell
C;Comment: The molecular mass is 23,500 and the pI is 4.8. The amino-terminus is bloc
C;Keywords: brain


  Query Match             31.1%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           |||
Db       5 LSG 7

RESULT  22
PT0557
T-cell receptor beta chain V-D-J region (126-1BD) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0557
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0557
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-8 <FEE>
A;Experimental source: day 18 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAD 6
           || |
Db       2 SGDD 5

RESULT  23
AKLQIM
locustamyoinhibiting peptide - migratory locust
C;Species: Locusta migratoria (migratory locust)
```

```
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 20-Mar-1998
C;Accession: A60065
R;Schoofs, L.; Holman, G.M.; Hayes, T.K.; Nachman, R.J.; De Loof, A.
Regul. Pept. 36, 111-119, 1991
A;Title: Isolation, identification and synthesis of locustamyoinhibiting peptide (LOM-MI
A;Reference number: A60065; MUID:92179466
A;Accession: A60065
A;Molecule type: protein
A;Residues: 1-9 <SCH>
C;Comment: This peptide hormone suppresses spontaneous contractions of the hindgut and-c
C;Superfamily: locustamyoinhibiting peptide
C;Keywords: amidated carboxyl end; hormone
F;9/Modified site: amidated carboxyl end (Trp) #status experimental


  Query Match             31.1%;  Score 14;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DIN 8
           |:|
Db       4 DLN 6


RESULT  24
A57444
neuropeptide Grb-AST B1 - two-spotted cricket
C;Species: Gryllus bimaculatus (two-spotted cricket)
C;Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 26-Jan-1996
C;Accession: A57444
R;Lorenz, M.W.; Kellner, R.; Hoffmann, K.H.
J. Biol. Chem. 270, 21103-21108, 1995
A;Title: A family of neuropeptides that inhibit juvenile hormone biosynthesis in the cri
A;Reference number: A57444; MUID:95403341
A;Accession: A57444
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <LOR>


  Query Match             31.1%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DIN 8
           |:|
Db       4 DLN 6


RESULT  25
B57444
neuropeptide Grb-AST B2 - two-spotted cricket
C;Species: Gryllus bimaculatus (two-spotted cricket)
C;Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 26-Jan-1996
C;Accession: B57444
R;Lorenz, M.W.; Kellner, R.; Hoffmann, K.H.
J. Biol. Chem. 270, 21103-21108, 1995
A;Title: A family of neuropeptides that inhibit juvenile hormone biosynthesis in the cri
A;Reference number: A57444; MUID:95403341
A;Accession: B57444
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <LOR>


  Query Match             31.1%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DIN 8
           |:|
Db       4 DLN 6
```

```
RESULT  26
C57444
neuropeptide Grb-AST B3 - two-spotted cricket
C;Species: Gryllus bimaculatus (two-spotted cricket)
C;Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 26-Jan-1996
C;Accession: C57444
R;Lorenz, M.W.; Kellner, R.; Hoffmann, K.H.
J. Biol. Chem. 270, 21103-21108, 1995
A;Title: A family of neuropeptides that inhibit juvenile hormone biosynthesis in the
A;Reference number: A57444; MUID:95403341
A;Accession: C57444
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <LOR>


  Query Match             31.1%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           |||
Db       5 LSG 7


RESULT  27
PT0268
Ig heavy chain CRD3 region (clone 3-94B) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C;Accession: PT0268
R;Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991
A;Title: Preferential utilization of specific immunoglobulin heavy chain diversity an
A;Reference number: PT0222; MUID:91108337
A;Accession: PT0268
A;Molecule type: DNA
A;Residues: 1-9 <YAM>
A;Experimental source: B lymphocyte
C;Keywords: heterotetramer; immunoglobulin


  Query Match             31.1%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   60.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 GADIN 8
           |  ||
Db       4 GIPIN 8


RESULT  28
S43959
Ig mu chain V region (clone 13) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 20-Oct-1994 #sequence_revision 26-Jul-1996 #text_change 17-Mar-1999
C;Accession: S43959
R;Wagner, S.D.; Williams, G.T.; Larson, T.; Neuberger, M.S.; Kitamura, D.; Rajewsky,
Nucleic Acids Res. 22, 1389-1393, 1994
A;Title: Antibodies generated from human immunoglobulin miniloci in transgenic mice.
A;Reference number: S43956; MUID:94248036
A;Accession: S43959
A;Molecule type: DNA
A;Residues: 1-4 <WAG>
C;Keywords: immunoglobulin


  Query Match             28.9%;  Score 13;  DB 2;  Length 4;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
```

```
          | :|
Db      1 YCAG 4


RESULT  29
I50210
gene c-rel protein - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 28-Feb-1997
C;Accession: I50210
R;Kabrun, N.; Bumstead, N.; Hayman, M.J.; Enrietto, P.J.
Mol. Cell. Biol. 10, 4788-4794, 1990
A;Title: Characterization of a novel promoter insertion in the c-rel locus.
A;Reference number: I50210; MUID:90355995
A;Accession: I50210
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-7 <KAB>
A;Cross-references: GB:M55577; NID:g555438; PID:g211661
C;Genetics:
A;Gene: c-rel


  Query Match             28.9%;  Score 13;  DB 2;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 LSGA 5
          ::||
Db      1 MAGA 4


RESULT  30
A58718
carnocin UI49 - Carnobacterium sp. (fragment)
C;Species: Carnobacterium sp.
C;Date: 23-Jan-1998 #sequence_revision 30-Jan-1998 #text_change 30-Jan-1998
C;Accession: A58718
R;Stoffels, G.; Nissen-Meyer, J.; Gudmundsdottir, A.; Sletten, K.; Holo, H.; Nes, I.F.
Appl. Environ. Microbiol. 58, 1417-1422, 1992
A;Title: Purification and characterization of a new bacteriocin isolated from a Carnobac
A;Reference number: A58718; MUID:92321768
A;Accession: A58718
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-7 <STO>
C;Keywords: antibiotic; lanthionine


  Query Match             28.9%;  Score 13;  DB 2;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 GADI 7
          |::|
Db      1 GSEI 4


RESULT  31
A41117
acetylcholinesterase (EC 3.1.1.7), venom - Asian cobra (fragment)
C;Species: Naja naja oxiana (Asian cobra, Oxus cobra)
C;Date: 27-Mar-1992 #sequence_revision 27-Mar-1992 #text_change 23-Jun-1993
C;Accession: A41117
R;Kreienkamp, H.J.; Weise, C.; Raba, R.; Aaviksaar, A.; Hucho, F.
Proc. Natl. Acad. Sci. U.S.A. 88, 6117-6121, 1991
A;Title: Anionic subsites of the catalytic center of acetylcholinesterase from Torpedo a
A;Reference number: A41117; MUID:91296772
A;Accession: A41117
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-8 <KRE>
C;Keywords: carboxylic ester hydrolase
```

```
  Query Match             28.9%;  Score 13;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 GADI 7
          ||::
Db      1 GAEM 4


RESULT  32
PT0547
T-cell receptor beta chain V-D-J region (126-1AI) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0547
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0547
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-8 <FEE>
A;Experimental source: day 18 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             28.9%;  Score 13;  DB 2;  Length 8;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 SGAD 6
          | ||
Db      3 SDAD 6


RESULT  33
QDRB
delta sleep-inducing peptide - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 20-Jun-2000 #sequence_revision 20-Jun-2000 #text_change 20-Jun-2000
C;Accession: A01422
R;Monnier, M.; Dudler, L.; Gachter, R.; Maier, P.F.; Tobler, H.J.; Schoenenberger, G.
Experientia 33, 548-552, 1977
A;Title: The delta sleep inducing peptide (DSIP). Comparative properties of the origi
A;Reference number: A01422; MUID:77185324
A;Accession: A01422
A;Molecule type: protein
A;Residues: 1-9 <MON>
C;Comment: This peptide was obtained from dialysates of occipital venous sinus blood
 of recipient rabbits, it induces spindle and delta EEG activity and reduced motor ac
C;Superfamily: unassigned animal peptides


  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 SGAD 6
          :| |
Db      2 AGGD 5


RESULT  34
C41170
photosystem II protein psbM - Chlamydomonas reinhardtii (fragment)
C;Species: Chlamydomonas reinhardtii
C;Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 30-Sep-1993
C;Accession: C41170
R;de Vitry, C.; Diner, B.A.; Popot, J.L.
J. Biol. Chem. 266, 16614-16621, 1991
```

```
A;Title: Photosystem II particles from Chlamydomonas reinhardtii. Purification, molecula
A;Reference number: A41170; MUID:91358452
A;Accession: C41170
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <DE5>

  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSGA 5
            ::||
Db        4 IAGA 7


RESULT  35
A61386
macrophage inhibitory factor (F5 cells) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 07-May-1999
C;Accession: A61386
R;Oki, S.; Hirose, S.I.; Higuchi, M.; Osawa, T.
Lymphokine Cytokine Res. 10, 273-280, 1991
A;Title: Macrophage migration inhibitory factor (MIF) produced by a human T cell hybrid
A;Reference number: A61386; MUID:92032107
A;Accession: A61386
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <OKI>


  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 ADI 7
            ||:
Db        2 ADV 4


Search completed: December 16, 2000, 03:35:12
Job time: 5645 sec
```

```
                    GenCore version 4.5
             Copyright (c) 1993 - 2000  Compugen Ltd.


OM protein - protein search, using sw model

Run on:          December 16, 2000, 03:07:18 ; Search time 107.12 Seconds
                                              (without alignments)
                                              2.873 Million cell updates/sec

Title:           US-09-529-121-3
Perfect score:   45
Sequence:        1 YLSGADINL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters:      61695

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       A_Geneseq_36:*
                1:  /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
                2:  /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
                3:  /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
                4:  /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
                5:  /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
                6:  /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
                7:  /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
                8:  /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
                9:  /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
                10:  /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
                11:  /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
                12:  /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
                13:  /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
                14:  /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
                15:  /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
                16:  /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
                17:  /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
                18:  /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
                19:  /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
                20:  /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
                21:  /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                   %
Result           Query
  No.    Score   Match Length DB  ID                      Description
---------------------------------------------------------------------------
    1       45   100.0      9  20  Y09527                  Carcinoembryonic a
    2       43    95.6      9  20  Y09526                  Carcinoembryonic a
    3       40    88.9      9  20  Y09528                  Carcinoembryonic a
    4       38    84.4      9  18  W39723                  Human carcina-embr
    5       38    84.4      9  19  W77134                  CEA synthetic pept
    6       38    84.4      9  19  W70045                  CEA derived HLA-A2
    7       38    84.4      9  20  Y47655                  Immunogenic peptid
    8       38    84.4      9  20  Y09525                  Carcinoembryonic a
    9       35    77.8      9  21  Y54173                  HLA binding peptid
   10       34    75.6      9  20  Y09529                  Carcinoembryonic a
   11       31    68.9      9  17  W00680                  Peptide comprising
   12       25    55.6      7  20  Y41846                  Rheumatoid arthrit
```

```
    13      24    53.3      9  19  W70078                  B. stearothermophi
    14      23    51.1      7  20  Y41847                  Rheumatoid arthrit
    15      23    51.1      9  11  R07966                  Tryptic fragment T
    16      23    51.1      9  18  W38383                  Synthetic pMEL17 p
    17      23    51.1      9  20  Y47062                  Immunogenic peptid
    18      22    48.9      9  16  R67605                  Jojoba fatty acyl-
    19      22    48.9      9  16  R67613                  Jojoba fatty acyl-
    20      22    48.9      9  17  W00690                  NCA analogue of re
    21      22    48.9      9  21  Y70832                  BH3 domain of mous
    22      21    46.7      6  11  R09414                  LFA-1 alpha subuni
    23      21    46.7      6  17  R88476                  Internal tryptic p
    24      21    46.7      9  19  W54298                  Human cytohesin-1
    25      20    44.4      7  20  Y41962                  Rheumatoid arthrit
    26      20    44.4      7  20  Y42028                  Rheumatoid arthrit
    27      20    44.4      9  19  W70077                  Thermus thermophil
    28      20    44.4      9  20  Y55545                  HLA binding plu-1
    29      20    44.4      9  20  Y55587                  HLA binding plu-1
    30      20    44.4      9  20  Y47818                  Immunogenic peptid
    31      20    44.4      9  20  Y05407                  Human BAK BH3 doma
    32      20    44.4      9  21  Y70831                  BH3 domain of huma
    33      20    44.4      9  21  Y54210                  HLA binding peptid
    34      19    42.2      6  17  W02264                  Gingivalis adhesio
    35      19    42.2      6  19  W56879                  Enzyme inhibitor p
    36      19    42.2      8  13  R26159                  Recombinant signal
    37      19    42.2      8  19  W56983                  Enzyme inhibitor p
    38      19    42.2      9  14  R30155                  MAb GAH variable r
    39      19    42.2      9  16  R87430                  Human MHC class II
    40      19    42.2      9  16  R76525                  MTAse peptide for
    41      19    42.2      9  19  W54299                  Human cytohesin-1
    42      19    42.2      9  19  W40267                  K. oxytoca R-speci
    43      19    42.2      9  20  Y30946                  Human antigenic MT
    44      19    42.2      9  21  Y54211                  HLA binding peptid
    45      19    42.2      9  21  Y56575                  Tyrosinase 2 deriv
    46      18    40.0      5  20  Y28182                  Factor Xa inhibiti
    47      18    40.0      6  17  R90466                  Antibody 561 displ
    48      18    40.0      6  19  W83890                  Peptide specific a
    49      18    40.0      6  20  Y55201                  Anti CD34 antibody
    50      18    40.0      6  20  Y26120                  Phospholipase A2 i
    51      18    40.0      6  21  Y86947                  Human haematopoiet
    52      18    40.0      7  18  W12026                  Herpes virus DNA p
    53      18    40.0      7  19  W69333                  Haemoglobin mutant
    54      18    40.0      7  20  Y17023                  Heat shock protein
    55      18    40.0      8  10  P91779                  Synthetic SCM-acti
    56      18    40.0      8  16  R76988                  Immune stimulating
    57      18    40.0      8  16  R73336                  Human TSH receptor
    58      18    40.0      8  17  W07241                  HIV-1 group O stra
    59      18    40.0      8  17  W07490                  Soybean pepsin dig
    60      18    40.0      8  19  W64478                  Human eosinophil c
    61      18    40.0      8  19  W53313                  CS4-CFA/I family s
    62      18    40.0      8  20  Y16852                  Heat shock protein
    63      18    40.0      8  20  W94102                  VIF-derived HIV pr
    64      18    40.0      8  20  W82429                  PVX-2 viral replic
    65      18    40.0      9  13  R20729                  Pancreas-elastase-
    66      18    40.0      9  14  R46546                  Jardetzky self pep
    67      18    40.0      9  15  Y38246                  HIV-derived HLA-bi
    68      18    40.0      9  15  R59233                  Peptide fragment (
    69      18    40.0      9  16  R70067                  Control peptide 92
    70      18    40.0      9  17  W49278                  Human leucocyte an
    71      18    40.0      9  17  R99964                  Clah12 B-cell epit
    72      18    40.0      9  19  W54515                  Synthetic polypept
    73      18    40.0      9  20  Y41907                  Rheumatoid arthrit
    74      18    40.0      9  20  Y42029                  Rheumatoid arthrit
    75      18    40.0      9  21  Y51905                  Elastase immunogen

                                   ALIGNMENTS


RESULT   1
Y09527
ID   Y09527 standard; peptide; 9 AA.
XX
AC   Y09527;
```

```
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:3.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;


  Query Match             100.0%;  Score 45;  DB 20;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
             |||||||||
Db       1 ylsgadinl 9


RESULT   2
Y09526
ID   Y09526 standard; peptide; 9 AA.
XX
AC   Y09526;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:2.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;


  Query Match             95.6%;  Score 43;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
             ||||||:||
Db       1 ylsgadlnl 9


RESULT   3
Y09528
ID   Y09528 standard; peptide; 9 AA.
XX
AC   Y09528;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:4.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
```

```
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;


  Query Match             88.9%;  Score 40;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
           |||||:|||
Db       1 ylsganinl 9


RESULT   4
W39723
ID   W39723 standard; peptide; 9 AA.
XX
AC   W39723;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Human carcina-embryonic antigen (CEA) peptide (pos. 571-579).
XX
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
XX
OS   Homo sapiens.
XX
PN   WO9741440-A1.
XX
PD   06-NOV-1997.
XX
PF   28-APR-1997;   97WO-NL00229.
XX
PR   23-DEC-1996;   96EP-0203670.
PR   26-APR-1996;   96EP-0201145.
XX
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
PI   Kast WM,  Melief CJM,  Offringa R,  Toes REM,  Van Der Burg SH;
XX
DR   WPI; 1997-549891/50.
XX
PT   Method of selecting T cell peptide epitope(s) - by measuring the
PT   stability of HLA class I-peptide complexes on intact B cells
XX
PS   Example 3; Page 85; 109pp; English.
XX
CC   Peptides W39430-W39734 are used in a novel method for the selection of
```

```
CC   immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC   method involves the identification of peptide sequences capable of
CC   binding to an HLA (human leukocyte antigen) class I molecule and
CC   measuring the binding of this epitope peptide to the HLA class I peptide.
CC   The stability of binding of the peptide and MHC (major histocompatablilty
CC   complex) class I molecule is measured on intact human B cells carrying
CC   the MHC molecule at their cell surfaces. The method can be used to select
CC   peptide epitopes for generating vaccines against a disease associated
CC   with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC   especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC   immune responses. Peptide W39723 is derived from the human
CC   carcino-embryonic antigen (CEA) and has the ability to bind to the human
CC   MHC Class I allele HLA-A2.1.
XX
SQ   Sequence   9 AA;


  Query Match             84.4%;  Score 38;  DB 18;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
           |||||::||
Db       1 ylsganlnl 9


RESULT   5
W77134
ID   W77134 standard; peptide; 9 AA.
XX
AC   W77134;
XX
DT   16-NOV-1998  (first entry)
XX
DE   CEA synthetic peptide epitope 1.
XX
KW   Tyrosinase; tyrosinase cytotoxic lymphocyte response;
KW   cytotoxic T lymphocyte; cysteine-depleted; melanoma.
XX
OS   Synthetic.
XX
PN   WO9833810-A2.
XX
PD   06-AUG-1998.
XX
PF   29-JAN-1998;   98WO-US01592.
XX
PR   30-JAN-1997;   97US-0037781.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI   Engelhard VH,  Hunt DF,  Kittlesen D,  Slingluff CL;
XX
DR   WPI; 1998-437388/37.
XX
PT   Disease specific immunogen  - comprises disease specific cytotoxic T
PT   lymphocyte epitope used to elicit melanoma specific CTL response
XX
PS   Disclosure; Page 27; 93pp; English.
XX
CC   The peptide epitope W77119-W77138 were created for human tumour-specific
CC   cytotoxic T lymphocyte response.  These peptides are are cysteine-
CC   depleted mutants of a native disease-specific CTL epitope.  The cysteine-
CC   depleted CTL epitopes elicit a stronger or more specific CTL response
CC   than the native epitope.  The epitopes can be used in a disease-specific
CC   immunogen to protect a mammal against disease in particular melanomas.
CC   The peptides may also be used to screen a sample for the presence of
CC   an antigen with the same epitope, or with a different cross-reactive
CC   epitope.
XX
SQ   Sequence   9 AA;
```

```
  Query Match             84.4%;  Score 38;  DB 19;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              |||||::||
Db          1 ylsganlnl 9

RESULT   6
W70045
ID   W70045 standard; peptide; 9 AA.
XX
AC   W70045;
XX
DT   22-OCT-1998  (first entry)
XX
DE   CEA derived HLA-A2.1 binding peptide 2 (residues 605-613).
XX
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW   antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW   viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW   fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9833888-A1.
XX
PD   06-AUG-1998.
XX
PF   30-JAN-1998;   98WO-US01959.
XX
PR   31-JAN-1997;   97US-0036696.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Celis E,  Sette A,  Sidney J,  Southwood S,  Tsai V;
XX
DR   WPI; 1998-437445/37.
XX
PT   Production of antigen-specific cytotoxic T cells - by incubating
PT   immunogenic peptide(s) from antigen that binds class I major
PT   histocompatibility complex molecules with pre-treated antigen
PT   presenting cells
XX
PS   Example 6; Page 75; 104pp; English.
XX
CC   Sequences shown in W70044 to W70052 represent peptides derived from
CC   carcinoembryonic antigen (CEA). The peptides can bind to a human
CC   leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method
CC   of invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC   vitro. The method comprises contacting immunogenic peptides from an
CC   antigen that binds class I major histocompatibility complex (MHC)
CC   molecules with antigen presenting cells (APCs) pretreated with
CC   pretreatment growth factors, and incubating the APCs with purified CD8
CC   cells in the presence of at least 2 incubation growth factors, thereby
CC   producing antigen-specific CTLs. A method for specifically killing
CC   target cells in a human patient is also provided which comprises
CC   obtaining a fluid sample containing CTLs from a patient, contacting the
CC   cytotoxic T cells with APCs pretreated with pre-treatment growth factors,
CC   where the APCs comprise class I MHC molecules. The pretreated APCs are
CC   incubated with the cytotoxic growth factors, thereby producing activated
CC   CTLs which are contacted with a carrier to form a composition. The
CC   composition can then be administered to the patient. The activated CTLs
CC   can be used for treating cancers, immune disorders, viral infections,
CC   AIDS, hepatitis, bacterial infection, fungal infection, malaria or
CC   tuberculosis.
XX
SQ   Sequence   9 AA;
```

```
  Query Match             84.4%;  Score 38;  DB 19;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADINL 9
              |||||::||
Db          1 ylsganlnl 9

RESULT   7
Y47655
ID   Y47655 standard; Peptide; 9 AA.
XX
AC   Y47655;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2266.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 118; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;

  Query Match             84.4%;  Score 38;  DB 20;  Length 9;
```

```
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADINL 9
            |||||::||
Db        1 ylsganlnl 9


RESULT   8
Y09525
ID   Y09525 standard; peptide; 9 AA.
XX
AC   Y09525;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist CAP-1.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 1; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects; for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             84.4%;  Score 38;  DB 20;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADINL 9
            |||||::||
Db        1 ylsganlnl 9


RESULT   9
Y54173
ID   Y54173 standard; peptide; 9 AA.
```

```
XX
AC   Y54173;
XX
DT   06-APR-2000  (first entry)
XX
DE   HLA binding peptide 1233.11 derived from source CEA.605V9.
XX
KW   Allele-specific binding motif; major histocompatibility complex; MHC;
KW   HLA; HLA-A2.1; cytotoxic T cell response; antigen; microbial infection;
KW   hepatitis; AIDS; malaria; condyloma acuminatum; cancer; prostate cancer;
KW   renal carcinoma; cervical carcinoma; lymphoma; tumour.
XX
OS   Unidentified.
XX
PN   WO9965522-A1.
XX
PD   23-DEC-1999.
XX
PF   17-JUN-1999;   99WO-US13789.
XX
PR   17-JUN-1998;   98US-0098584.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Sidney J,  Southwood S;
XX
DR   WPI; 2000-106018/09.
XX
PT   Novel HLA binding immunogenic peptides used to induce T cell activation
PT   and to induce an immune response  -
XX
PS   Claim 1; Page 32; 42pp; English.
XX
CC   Peptides Y54171-Y54236 represent immunogenic peptides comprising an
CC   allele-specific binding motif for the major histocompatibility complex
CC   (MHC) molecule HLA, e.g. HLA-A2.1. The peptides have conserved residues
CC   at certain positions such as positions 2 and 9. Also, the peptides do not
CC   comprise negative binding residues at other positions, such as positions
CC   1, 3, 6 and/or 7 (peptides 9 amino acids long) and at poisons 1, 3, 4,
CC   5, 7, 8 and/or 9 (peptides 10 amino acids long). The peptides are used to
CC   induce a cytotoxic T cell response to a preselected antigen. The method
CC   comprises contacting cytotoxic T cells from a patient (optionally
CC   expressing a specific MHC class I allele) with the present peptides.
CC   The peptides are used to treat and prevent microbial infection (e.g. in
CC   viral hepatitis B and C, human papilomavirus (HPV) infection, AIDS,
CC   cytomegalovirus (CMV), malaria, and condyloma acuminatum) and cancer
CC   (e.g. prostate cancer, renal carcinoma, cervical carcinoma, lymphoma).
CC   Patients in the acute phase of infection can be treated with the
CC   peptides in conjunction with other treatments. The antigenic peptides
CC   may be used to elicit cytotoxic T lymphocytes (CTLs) ex vivo and in
CC   vivo. The resulting CTLs can be used to treat chronic infections (viral
CC   or bacterial) or tumours in patients that do not respond to conventional
CC   forms of therapy. The peptides may also be used to produce monoclonal
CC   antibodies, which are useful as potential diagnostic or therapeutic
CC   agents. The peptides may also be used as diagnostic reagents.
XX
SQ   Sequence   9 AA;

  Query Match             77.8%;  Score 35;  DB 21;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADINL 9
            |||||::|:
Db        1 ylsganlnv 9


RESULT  10
Y09529
ID   Y09529 standard; peptide; 9 AA.
XX
AC   Y09529;
```

```
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:5.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             75.6%;  Score 34;  DB 20;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGADINL 9
           ||||| :||
Db       1 ylsgaclnl 9


RESULT  11
W00680
ID   W00680 standard; peptide; 9 AA.
XX
AC   W00680;
XX
DT   01-MAY-1997  (first entry)
XX
DE   Peptide comprising residues 571-579 of Carcinoembryonic antigen.
XX
KW   Carcinoembryonic; antigen; human; cytotoxic T cell; pox virus;
KW   vector; epitope; determination; screening; tumour; treatment.
XX
OS   Homo sapiens.
XX
PN   WO9626271-A1.
XX
PD   29-AUG-1996.
XX
PF   13-FEB-1996;   96WO-US02156.
XX
PR   22-FEB-1995;   95US-0396385.
XX
PA   (THER-) THERION BIOLOGICS CORP.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Panicali D,  Schlom J,  Tsang KY;
XX
DR   WPI; 1996-402364/40.
XX
PT   Generation of human cytotoxic T-cells specific for CEA - useful in
PT   therapy, epitope mapping and drug screening
XX
PS   Claim 4; Page 57; 76pp; English.
XX
CC   Producing carcinoembryonic antigen (CEA) specific human cytotoxic T
CC   cells (CTC), comprises introducing a 1st pox virus vector, having
CC   at least 1 insertion site containing a DNA segment encoding a CEA
CC   peptide (i.e. the present peptide) to a host to stimulate CTC
CC   production, and at least 1 periodic interval after that, contacting
CC   the host with an additional antigen. The CEA specific CTC can be
CC   used to determine the CTC eliciting epitope of CEA, and to screen
CC   for compounds which enhance the ability of the antigen to create a
CC   CTC response. A host with a CEA expressing tumour can be treated by
CC   introducing the CTC to the host, and at least 1 periodic interval
CC   after that introducing a CEA peptide, i.e. the present peptide.
CC   The present peptide is positive for binding to HLA-A2, and scored
CC   561 and 806 in T2 cell binding assays, where the binding of an
CC   appropriate peptide results in the upregulation of surface HLA-A2
CC   on the T2 cells, which can be quantified via FACScan using an
CC   anti-HLA-A2 antibody (background 280 and 300).
XX
SQ   Sequence   9 AA;

  Query Match             68.9%;  Score 31;  DB 17;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 2.1e+05;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGADINL 9
           ||||::||
Db       2 lsganlnl 9


RESULT  12
Y41846
ID   Y41846 standard; Peptide; 7 AA.
XX
AC   Y41846;
XX
DT   09-DEC-1999  (first entry)
XX
DE   Rheumatoid arthritis diagnostic protein isoform 17 peptide #1.
XX
KW   Human; rheumatoid arthritis; RA; diagnosis; RPI; RADF; detection;
KW   rheumatoid arthritis diagnostic feature; ERPI; synovial fluid;
KW   rheumatoid arthritis diagnostic protein isoform; screening;
KW   expression reference protein isoform; prognosis.
XX
OS   Homo sapiens.
XX
PN   WO9947925-A2.
XX
PD   23-SEP-1999.
XX
PF   15-MAR-1999;   99WO-GB00763.
XX
PR   13-MAR-1998;   98GB-0005477.
XX
```

```
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Parekh RB,  Patel TP,  Townsend RR;
XX
DR   WPI; 1999-571871/48.
XX
PT   Diagnosis of human rheumatoid arthritis by two-dimensional
PT   electrophoresis -
XX
PS   Claim 20; Page 150; 157pp; English.
XX
CC   A method has been developed for the diagnosis of human rheumatoid
CC   arthritis (RA) using two-dimensional electrophoresis to generate a
CC   two-dimensional array of features. The method can be used for screening,
CC   diagnosis and prognosis of RA in a subject or for monitoring the effect
CC   of an anti-RA drug or therapy administered to a subject. The method
CC   comprises: (a) analysing a sample of serum or plasma and optionally
CC   synovial fluid by two-dimensional electrophoresis, to generate a two-
CC   dimensional array of features; (b) identifying at least one chosen
CC   feature whose relative abundance correlates with the presence or absence
CC   of RA; and (c) comparing the abundance of each chosen feature in the
CC   sample with the abundance of that chosen feature in serum or plasma from
CC   one or more persons without RA, where the relative abundance of the
CC   chosen feature or features in the sample indicates the presence or
CC   absence of RA in the subject. The method can also be used in clinical
CC   studies for testing drugs for therapy of RA, for purification of RA-
CC   diagnostic protein isoforms (RPIs), and for production of antibodies to
CC   RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify
CC   compounds that promote or inhibit their activity, which are then used as
CC   RA drugs. Nucleic acid encoding RADFs can be used in gene therapy
CC   protocols. Y41844 to Y42100 represent RPI peptides, Y42101 to Y42103
CC   represent expression reference protein isoform peptides and Z25066 to
CC   Z25068 represent degenerate probes for RPIs, which are all used in
CC   the exemplification of the present invention.
XX
SQ   Sequence   7 AA;

  Query Match             55.6%;  Score 25;  DB 20;  Length 7;
  Best Local Similarity   83.3%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGADIN 8
           |||||:
Db       2 sgadis 7


RESULT  13
W70078
ID   W70078 standard; peptide; 9 AA.
XX
AC   W70078;
XX
DT   28-OCT-1998  (first entry)
XX
DE   B. stearothermophilus methionyl-tRNA synthetase fragment.
XX
KW   Mycobacterium tuberculosis; mycobacterial; methionyl-tRNA synthetase;
KW   enzyme; antibiotic; tuberculosis; Bacillus stearothermophilus.
XX
OS   Bacillus stearothermophilus.
XX
PN   US5798240-A.
XX
PD   25-AUG-1998.
XX
PF   11-JAN-1996;   96US-0584226.
XX
PR   13-SEP-1994;   94US-0305766.
PR   11-JAN-1996;   96US-0584226.
XX
PA   (CUBI-) CUBIST PHARM INC.
XX
```

```
PI   Kim S,  Lee SH,  Martinis SA,  Sassanfar M,  Schimmel PR;
XX
DR   WPI; 1998-480383/41.
XX
PT   Recombinant genes encoding mycobacterial amino acyl tRNA synthetases
PT   - useful for recombinant production of the enzyme for use in
PT   screening of antibiotics against Mycobacterium tuberculosis
XX
PS   Example 1; Columns 33-34; 32pp; English.
XX
CC   Sequences shown in W70077 to W70086 represent methionyl-tRNA (Met-tRNA)
CC   synthetase fragments from different bacterial species. These are used
CC   for designing degenerate primers (V43887 to V43890) used for isolating
CC   Met-tRNa sythetase gene fragment by PCR from M. tuberculosis and M.
CC   kansasii. The invention provides an expression vector comprising a
CC   nucleic acid encoding a mycobacterial aminoacyl-tRNAs (atRNA) (especially
CC   Met-tRNA) synthetase under control of transcriptional signals that can be
CC   used to transform suitable host cells. The nucleict acids and host cells
CC   are used for the recombinant production of mycobacterial amino acyl tRNA
CC   synthetases, especially of methionyl tRNA synthetase. The enzymes are
CC   used by the organism in protein synthesis, and as such, the recombinant
CC   enzyme can be used to identify candidate drugs for use as antibiotics
CC   towards mycobacteria, especially M. tuberculosis, which is responsible
CC   for tuberculosis. Antisense constructs of the nucleic acids can also be
CC   used in antisense inhibition of the synthetase gene. The recombinant
CC   enzyme allows quick assays in screening of antibiotics. Present testing
CC   protocols involve exposing whole mycobacteria to candidate drugs and
CC   seeing their effect. This is time consuming as the bacteria are generally
CC   slow growing. Use of the enzyme also prevents researchers from having to
CC   work with pathogenic strains.
XX
SQ   Sequence   9 AA;

  Query Match             53.3%;  Score 24;  DB 19;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGAD 6
           ||:| |
Db       1 yltgtd 6


RESULT  14
Y41847
ID   Y41847 standard; Peptide; 7 AA.
XX
AC   Y41847;
XX
DT   09-DEC-1999  (first entry)
XX
DE   Rheumatoid arthritis diagnostic protein isoform 17 peptide #2.
XX
KW   Human; rheumatoid arthritis; RA; diagnosis; RPI; RADF; detection;
KW   rheumatoid arthritis diagnostic feature; ERPI; synovial fluid;
KW   rheumatoid arthritis diagnostic protein isoform; screening;
KW   expression reference protein isoform; prognosis.
XX
OS   Homo sapiens.
XX
PN   WO9947925-A2.
XX
PD   23-SEP-1999.
XX
PF   15-MAR-1999;   99WO-GB00763.
XX
PR   13-MAR-1998;   98GB-0005477.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Parekh RB,  Patel TP,  Townsend RR;
XX
DR   WPI; 1999-571871/48.
```

```
XX
PT   Diagnosis of human rheumatoid arthritis by two-dimensional
PT   electrophoresis -
XX
PS   Claim 20; Page 150; 157pp; English.
XX
CC   A method has been developed for the diagnosis of human rheumatoid
CC   arthritis (RA) using two-dimensional electrophoresis to generate a
CC   two-dimensional array of features. The method can be used for screening,
CC   diagnosis and prognosis of RA in a subject or for monitoring the effect
CC   of an anti-RA drug or therapy administered to a subject. The method
CC   comprises: (a) analysing a sample of serum or plasma and optionally
CC   synovial fluid by two-dimensional electrophoresis, to generate a two-
CC   dimensional array of features; (b) identifying at least one chosen
CC   feature whose relative abundance correlates with the presence or absence
CC   of RA; and (c) comparing the abundance of each chosen feature in the
CC   sample with the abundance of that chosen feature in serum or plasma from
CC   one or more persons without RA, where the relative abundance of the
CC   chosen feature or features in the sample indicates the presence or
CC   absence of RA in the subject. The method can also be used in clinical
CC   studies for testing drugs for therapy of RA, for purification of RA-
CC   diagnostic protein isoforms (RPIs), and for production of antibodies to
CC   RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify
CC   compounds that promote or inhibit their activity, which are then used as
CC   RA drugs. Nucleic acid encoding RADFs can be used in gene therapy
CC   protocols. Y41844 to Y42100 represent RPI peptides, Y42101 to Y42103
CC   represent expression reference protein isoform peptides and Z25066 to
CC   Z25068 represent degenerate probes for RPIs, which are all used in
CC   the exemplification of the present invention.
XX
SQ   Sequence   7 AA;


  Query Match             51.1%;  Score 23;  DB 20;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGADIN 8
           ||||::
Db       2 sgadls 7


RESULT  15
R07966
ID   R07966 standard; protein; 9 AA.
XX
AC   R07966;
XX
DT   14-JAN-1991  (first entry)
XX
DE   Tryptic fragment T32 of human N-lipocortin.
XX
KW   Human N-lipocortin; placenta; inflammation reduction; arthritis;
KW   phospholipase A2 inhibitor; tryptic fragment T32.
XX
OS   Homo sapiens.
XX
PN   US4950646-A.
XX
PD   21-AUG-1990.
XX
PF   10-JAN-1986;   86US-0929199.
XX
PR   10-JAN-1986;   86US-0929199.
PR   05-SEP-1985;   85US-0772892.
PR   14-AUG-1985;   85US-0765877.
PR   15-MAR-1985;   85US-0712376.
PR   10-JAN-1985;   85US-0690146.
XX
PA   (BIOJ ) BIOGEN NV.
XX
PI   Wallner BP,  Pepinsky RB,  Garwin JL,  Schindler DG,  Huang KS;
XX
DR   WPI; 1990-274549/36.
XX
PT   Pure fragment of human lipocortin - useful for reducing
PT   inflammation or for treating arthritis, etc.
XX
PS   Disclosure; Fig 25; 51pp; English.
XX
CC   T32 corresponds to a peak from the tryptic map of N-lipocortin,
CC   isolated from human placenta, on a Speed Vac Concentration.
CC   Amino acids 1 can also be G.
CC   Based on the similarity in the phospholipase A2 inhibitory activity
CC   of lipocortin and N-lipocortin and the similarity in the protein and DNA
CC   sequences, it was concluded that the two proteins represent a
CC   family of related proteins.  There is ca. 60% homology.
CC   The protein can be used for reducing inflammation or treating
CC   arthritic, allergic, dermatologic, opthalmic and collagen diseases
CC   and other diseases involving inflammation processes.
CC   See also Q05805-25, Q06581, R07926-37 and R07956-66.
XX
SQ   Sequence   9 AA;

  Query Match             51.1%;  Score 23;  DB 11;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGAD 6
           || | |
Db       4 ylxggd 9


RESULT  16
W38383
ID   W38383 standard; peptide; 9 AA.
XX
AC   W38383;
XX
DT   08-APR-1998  (first entry)
XX
DE   Synthetic pMEL17 peptide.
XX
KW   Melanoma; immunogen; cytotoxic T lymphocyte; CTL;
KW   human leukocyte antigen-A1; HLA-A1; human leukocyte antigen-A3;
KW   HLA-A3; epitope; pMEL-17; tyrosinase; vaccine; protection.
XX
OS   Synthetic.
XX
PN   WO9734613-A1.
XX
PD   25-SEP-1997.
XX
PF   17-MAR-1997;   97WO-US04958.
XX
PR   04-OCT-1996;   96US-0027627.
PR   19-MAR-1996;   96US-0013972.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI   Cox AL,  Engelhard VH,  Hendrikson RC,  Hunt DF,  Kittlesen D;
PI   Shabanowitz J,  Skipper J,  Slingluff CL;
XX
DR   WPI; 1997-479982/44.
XX
PT   Melanoma-specific immunogens of pMel-17 and tyrosinase - useful in
PT   vaccination for producing melanoma-specific cytotoxic T lymphocytes
XX
PS   Example 9; Page 65; 106pp; English.
XX
CC   The present peptide was used in the preparation of a novel melanoma
CC   specific immunogen, comprising at least 1 melanoma specific
CC   cytotoxic T lymphocyte (CTL) epitope, where at least 1 of the
CC   epitopes is substantially homologous to a human leukocyte
CC   antigen-A1 (HLA-A1) and HLA-A3 restricted epitope of a melanoma
```

```
CC   antigen, either pMEL-17 or tyrosinase. The immunogen can be used in
CC   vaccines for protection against melanoma in mammals.
XX
SQ   Sequence   9 AA;

  Query Match             51.1%;  Score 23;  DB 18;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADIN 8
              ||: ||::
Db          1 ylaeadls 8


RESULT  17
Y47062
ID   Y47062 standard; Peptide; 9 AA.
XX
AC   Y47062;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #1673.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 92; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;

  Query Match             51.1%;  Score 23;  DB 20;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADIN 8
              ||: ||::
Db          1 ylaeadls 8


RESULT  18
R67605
ID   R67605 standard; Protein; 9 AA.
XX
AC   R67605;
XX
DT   14-AUG-1995  (first entry)
XX
DE   Jojoba fatty acyl-CoA reductase 56kd protein fragment.
XX
KW   Jojoba; fatty acyl CoA reductase; fatty acyl; fatty alcohol.
XX
OS   Simmondsia chinensis.
XX
PN   US5370996-A.
XX
PD   06-DEC-1994.
XX
PF   22-FEB-1991;   91US-0659975.
XX
PR   22-FEB-1991;   91US-0659975.
PR   27-SEP-1991;   91US-0767251.
PR   20-NOV-1991;   91US-0796256.
PR   31-JUL-1992;   92US-0920430.
XX
PA   (CALJ ) CALGENE INC.
XX
PI   Lassner MW,  Metz JG,  Pollard MR;
XX
DR   WPI; 1995-021884/03.
XX
PT   New recombinant constructs for transforming plants of E. coli  -
PT   contg. nucleic acid encoding jojoba embryo long chain fatty
PT   acyl-CoA reductase.
XX
PS   Example 4; Column 26; 30pp; English.
XX
CC   The sequence encoding the jojoba embryo long chain fatty acyl-CoA
CC   reductase may be used in recombinant constructs which in turn can be
CC   used to transform E. coli.  Such constructs are useful for the
CC   expression of the jojoba embryo long chain fatty acyl CoA reductase
CC   in host cells.  The enzyme catalyses the formation of a fatty
CC   alcohol from a fatty acyl substrate.  Analysis of the enzyme by SDS
CC   PAGE gave two prominent bands having apparent molecular masses of
CC   approximately 52 and 54 kD.  As the apparent size of the reductase
CC   enzyme in the native state is approximately 49 kD as determined by
CC   size exclusion chromatography, these bands probably represented two
CC   related forms of the enzyme instead of two different subunits of the
CC   enzyme. This sequence is a peptide fragment from the 56 kD protein.
XX
SQ   Sequence   9 AA;

  Query Match             48.9%;  Score 22;  DB 16;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 GADINL 9
              | |||:
```

```
Db      2 gldinv 7


RESULT  19
R67613
ID   R67613 standard; Protein; 9 AA.
XX
AC   R67613;
XX
DT   14-AUG-1995  (first entry)
XX
DE   Jojoba fatty acyl-CoA reductase 54kd protein fragment.
XX
KW   Jojoba; fatty acyl CoA reductase; fatty acyl; fatty alcohol.
XX
OS   Simmondsia chinensis.
XX
PN   US5370996-A.
XX
PD   06-DEC-1994.
XX
PF   22-FEB-1991;   91US-0659975.
XX
PR   22-FEB-1991;   91US-0659975.
PR   27-SEP-1991;   91US-0767251.
PR   20-NOV-1991;   91US-0796256.
PR   31-JUL-1992;   92US-0920430.
XX
PA   (CALJ ) CALGENE INC.
XX
PI   Lassner MW,  Metz JG,  Pollard MR;
XX
DR   WPI; 1995-021884/03.
XX
PT   New recombinant constructs for transforming plants of E. coli  -
PT   contg. nucleic acid encoding jojoba embryo long chain fatty
PT   acyl-CoA reductase.
XX
PS   Example 4; Column 26; 30pp; English.
XX
CC   The sequence encoding the jojoba embryo long chain fatty acyl-CoA
CC   reductase may be used in recombinant constructs which in turn can be
CC   used to transform E. coli.  Such constructs are useful for the
CC   expression of the jojoba embryo long chain fatty acyl CoA reductase
CC   in host cells.  The enzyme catalyses the formation of a fatty
CC   alcohol from a fatty acyl substrate.  Analysis of the enzyme by SDS
CC   PAGE gave two prominent bands having apparent molecular masses of
CC   approximately 52 and 54 kD.  As the apparent size of the reductase
CC   enzyme in the native state is approximately 49 kD as determined by
CC   size exclusion chromatography, these bands probably represented two
CC   related forms of the enzyme instead of two different subunits of the
CC   enzyme. This sequence is a peptide fragment from the 54 kD protein.
XX
SQ   Sequence   9 AA;


  Query Match             48.9%;  Score 22;  DB 16;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 GADINL 9
          | |||:
Db      2 gldinv 7


RESULT  20
W00690
ID   W00690 standard; peptide; 9 AA.
XX
AC   W00690;
XX
DT   01-MAY-1997  (first entry)
XX
```

```
DE   NCA analogue of residues 571-579 of carcinoembryonic antigen.
XX
KW   Carcinoembryonic; antigen; epitope; NCA; analogue.
XX
OS   Homo sapiens.
XX
PN   WO9626271-A1.
XX
PD   29-AUG-1996.
XX
PF   13-FEB-1996;   96WO-US02156.
XX
PR   22-FEB-1995;   95US-0396385.
XX
PA   (THER-) THERION BIOLOGICS CORP.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Panicali D,  Schlom J,  Tsang KY;
XX
DR   WPI; 1996-402364/40.
XX
PT   Generation of human cytotoxic T-cells specific for CEA - useful in
PT   therapy, epitope mapping and drug screening
XX
PS   Example 2; Page 60; 76pp; English.
XX
CC   The present peptide is negative for binding to HLA-A2, and scored
CC   252 and 225 in T2 cell binding assays, where the binding of an
CC   appropriate peptide results in the upregulation of surface HLA-A2
CC   on the T2 cells, which can be quantified via FACScan using an
CC   anti-HLA-A2 antibody (background 280 and 300).
XX
SQ   Sequence   9 AA;


  Query Match             48.9%;  Score 22;  DB 17;  Length 9;
  Best Local Similarity   44.4%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 YLSGADINL 9
          |  | ::||
Db      1 yrpgenlnl 9


RESULT  21
Y70832
ID   Y70832 standard; peptide; 9 AA.
XX
AC   Y70832;
XX
DT   31-JUL-2000  (first entry)
XX
DE   BH3 domain of mouse BAK protein.
XX
KW   Mouse; truncated BAX protein; tBAX; BAX alpha; BCL-2 family; head trauma;
KW   neuron; anti-apoptotic; cerebroprotective; neuroprotective; neuroactive;
KW   apoptosis; treatment; neurodegenerative disease; peripheral nerve injury;
KW   spinal cord injury; stroke; pro-apoptotic; PCD; programmed cell death;
KW   BH3 domain; BCL-2 homology domain; BAK protein.
XX
OS   Mus musculus.
XX
PN   WO200023083-A1.
XX
PD   27-APR-2000.
XX
PF   22-OCT-1999;   99WO-US24747.
XX
PR   22-OCT-1998;   98US-0177315.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Johnson EM,  Easton R;
```

```
XX
DR   WPI; 2000-339513/29.
XX
PT   Truncated BAX polypeptides useful for preventing apoptosis of neurons
PT   for the treatment of nervous system disorders -
XX
PS   Disclosure; Fig 3; 43pp; English.
XX
CC   The patent discloses specific truncated BAX (tBAX) proteins
CC   which inhibit neuronal apoptosis induced by trophic factor deprivation.
CC   The anti-apoptotic tBAX proteins include tBAX70, tBAX78
CC   and their mutants. These proteins contain the N-terminal region and at
CC   least a portion of the BH3 domain of BAX alpha and lack the BH1, BH2 and
CC   C-terminal transmembrane domains. The tBAX protein lacking only the
CC   transmembrane domain has been shown to have anti-apoptotic activity.
CC   The tBAX proteins are used to treat diseases associated with neuronal
CC   apoptosis, e.g. neurodegenerative diseases, peripheral nerve injury,
CC   spinal cord injury, head trauma and stroke. tBAX proteins
CC   truncating at amino acid 68 of the BH3 domain of BAX alpha also have
CC   anti-apoptotic activity because Asp at position 68 is shown to be
CC   important for BAX activity and is conserved in all BCL-2 family members.
CC   The present sequence is a BH3 domain of mouse BAK protein, a
CC   pro-apoptotic protein belonging to BCL-2 family that is involved in
CC   regulation of neuronal programmed cell death. The present sequence
CC   contains the conserved Asp residue.
XX
SQ   Sequence   9 AA;

  Query Match             48.9%;  Score 22;  DB 21;  Length 9;
  Best Local Similarity   71.4%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        2 LSGADIN 8
            | | |||
Db        3 ligddin 9


RESULT  22
R09414
ID   R09414 standard; Peptide; 6 AA.
XX
AC   R09414;
XX
DT   23-AUG-1990  (first entry)
XX
DE   LFA-1 alpha subunit polypeptide (k).
XX
KW   Lymphocyte function associated antigen; inflammation; metastasis.
XX
PN   EP362526-A.
XX
PD   11-APR-1990.
XX
PF   17-AUG-1989;   89EP-0115160.
XX
PR   23-AUG-1988;   88US-0235227.
PR   09-MAR-1989;   89US-0321017.
XX
PA   (DANA-) DANA FARBER CANCER.
XX
PI   Springer TA,  Larson R;
XX
DR   WPI; 1990-108985/15.
XX
PT   Pure alpha subunit of lymphocyte function associated antigen -
PT   and encoding DNA sequences, useful eg for suppressing
PT   inflammation or metastasis
XX
PS   Claim 6; Page 19; 27pp; English.
XX
CC   The alpha-subunit (a-SU), contg. al least one of the polypeptides
CC   given in R09404-417, can bind to ICAM-1 (or other natural
```

```
CC   ligands) on the surface of cells, and can associate with the beta-SU
CC   to form a heterodimer (also able to bind to ICAM-I).  a-SU, and its
CC   derivs., are useful in suppressing inflammation, metastasis and
CC   growth of a-SU expressing tumour cells and is used in the treatment
CC   of viral infections.
CC   The pref. dose is 1 pg - 10 mg/kg.
XX
SQ   Sequence   6 AA;

  Query Match             46.7%;  Score 21;  DB 11;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSG 4
            ||||
Db        2 ylsg 5


RESULT  23
R88476
ID   R88476 standard; peptide; 6 AA.
XX
AC   R88476;
XX
DT   30-AUG-1996  (first entry)
XX
DE   Internal tryptic peptide from Tre6P synthase (peak 29) #2.
XX
KW   Tryptic peptide; trehalose-6-phosphate synthase; M. smegmatis; probe;
KW   trehalose; transgenic plant; heparin-activated; preservation; food;
KW   antigenic determinant; yeast; TSP1; fruit; berry; puree; jelly; jam.
XX
OS   Mycobacterium smegmatis.
XX
PN   WO9600789-A1.
XX
PD   11-JAN-1996.
XX
PF   29-JUN-1995;   95WO-FI00377.
XX
PR   29-JUN-1994;   94FI-0003133.
XX
PA   (ALKO-) ALKO GROUP LTD.
XX
PI   Holmstrom K,  Londesborough J,  Mandal A,  Mantyla E;
PI   Palva ET,  Tunnela O, Welin B;
XX
DR   WPI; 1996-077499/08.
XX
PT   New transgenic plants with increase trehalose contents - prepd. by
PT   transforming plants with a trehalose-6-phosphate synthase gene fused
PT   to a non-constitutive promoter
XX
PS   Example 6; Page 36; 55pp; English.
XX
CC   The sequences given in R88473-80 are internal tryptic peptides
CC   derived from trehalose-6-phosphate (Tre6P) synthase from M.
CC   smegmatis.  Tre6P is the key enzyme in the synthesis of trehalose
CC   via Tre6P.  The aim of the invention is to produce a transgenic
CC   plant with increase trehalose content.  Tre6P in M. smegmatis is
CC   heparin-activated and was isolated  and purified.    These peptides
CC   were derived from a protein which was purified with a mol. wt. of 55
CC   kD which  shared antigenic determinants with the yeast Tre6P
CC   synthase protein.  Using these peptides probes may be designed for
CC   the isolation of the Tre6P gene (TSP1) for the production of the
CC   transgenic plants.  The trehalose may be isolated from the transgenic
CC   plants and used in bulk preparation including the preservation of the
CC   flavour and structure of food stuffs during drying.  Fruits and berries
CC   form the transformed plants may be processed into purees, jellies and
CC   jams which have a fresher and richer flavour due to the increased
CC   trehalose content.
XX
```

```
SQ   Sequence   6 AA;

  Query Match             46.7%;  Score 21;  DB 17;  Length 6;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGA 5
            || ||
Db        1 ylega 5


RESULT  24
W54298
ID   W54298 standard; Peptide; 9 AA.
XX
AC   W54298;
XX
DT   28-AUG-1998  (first entry)
XX
DE   Human cytohesin-1 PH domain randomised AB loop (clone 7).
XX
KW   Pleckstrin homology domain; PH domain; peptide library;
KW   functional molecular surface; protein structural template;
KW   vaccine; gene therapy; cytohesin 1; human.
XX
OS   Synthetic.
XX
PN   WO9745538-A1.
XX
PD   04-DEC-1997.
XX
PF   30-MAY-1997;   97WO-EP02840.
XX
PR   31-MAY-1996;   96EP-0108776.
XX
PA   (MEDI-) MEDIGENE AG.
XX
PI   Bruhn H,  Funk M,  Henkel T,  Steipe B;
XX
DR   WPI; 1998-230215/20.
DR   N-PSDB; V26501.
XX
PT   Vectors used to produce PH domain-like peptide libraries - which are
PT   screened for therapeutically useful peptide(s), e.g. to produce
PT   vaccines
XX
PS   Example 5; Fig 12; 137pp; English.
XX
CC   This is a randomised AB loop encoded by clone 7 (see V26501) of a
CC   large peptide library (2 million clones) of synthetic Pleckstrin
CC   homology (PH) domains with randomised discontinuous surface
CC   epitopes.  The progenitor AB sequence comprises amino acid residues
CC   16-21 of a stabilised synthetic PH domain (see W54310) of human
CC   cytohesin 1.  The peptide library was generated by randomisation of
CC   AB and CD loop regions using randomised oligonucleotide primers
CC   (see W26493-94).  The randomised AB and CD loop regions of 9 clones
CC   (see W54292-309) of the peptide library are provided.  Randomisation
CC   did not compromise the structural integrity and folding stability of
CC   the progenitor domain.  The invention provides vectors that are used
CC   for the production of PH domain-like peptide libraries, which can
CC   be screened to identify peptides that have desirable properties,
CC   especially novel binding or catalytic properties, and which may
CC   be of use in research or therapy, or as vaccines.  Novel synthetic
CC   protein structural templates for the generation, screening and
CC   evolution of functional molecular surfaces are provided.
XX
SQ   Sequence   9 AA;

  Query Match             46.7%;  Score 21;  DB 19;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy        2 LSGADIN 8
            |||  :|
Db        1 lsggrvn 7


RESULT  25
Y41962
ID   Y41962 standard; Peptide; 7 AA.
XX
AC   Y41962;
XX
DT   09-DEC-1999  (first entry)
XX
DE   Rheumatoid arthritis diagnostic protein isoform peptide #113.
XX
KW   Human; rheumatoid arthritis; RA; diagnosis; RPI; RADF; detection;
KW   rheumatoid arthritis diagnostic feature; ERPI; synovial fluid;
KW   rheumatoid arthritis diagnostic protein isoform; screening;
KW   expression reference protein isoform; prognosis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 6
FT                   /label= Ile, Leu
XX
PN   WO9947925-A2.
XX
PD   23-SEP-1999.
XX
PF   15-MAR-1999;   99WO-GB00763.
XX
PR   13-MAR-1998;   98GB-0005477.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Parekh RB,  Patel TP,  Townsend RR;
XX
DR   WPI; 1999-571871/48.
XX
PT   Diagnosis of human rheumatoid arthritis by two-dimensional
PT   electrophoresis -
XX
PS   Disclosure; Page 20; 157pp; English.
XX
CC   A method has been developed for the diagnosis of human rheumatoid
CC   arthritis (RA) using two-dimensional electrophoresis to generate a
CC   two-dimensional array of features. The method can be used for screening,
CC   diagnosis and prognosis of RA in a subject or for monitoring the effect
CC   of an anti-RA drug or therapy administered to a subject. The method
CC   comprises: (a) analysing a sample of serum or plasma and optionally
CC   synovial fluid by two-dimensional electrophoresis, to generate a two-
CC   dimensional array of features; (b) identifying at least one chosen
CC   feature whose relative abundance correlates with the presence or absence
CC   of RA; and (c) comparing the abundance of each chosen feature in the
CC   sample with the abundance of that chosen feature in serum or plasma from
CC   one or more persons without RA, where the relative abundance of the
CC   chosen feature or features in the sample indicates the presence or
CC   absence of RA in the subject. The method can also be used in clinical
CC   studies for testing drugs for therapy of RA, for purification of RA-
CC   diagnostic protein isoforms (RPIs), and for production of antibodies to
CC   RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify
CC   compounds that promote or inhibit their activity, which are then used as
CC   RA drugs. Nucleic acid encoding RADFs can be used in gene therapy
CC   protocols. Y41844 to Y42100 represent RPI peptides, Y42101 to Y42103
CC   represent expression reference protein isoform peptides and Z25066 to
CC   Z25068 represent degenerate probes for RPIs, which are all used in
CC   the exemplification of the present invention.
XX
SQ   Sequence   7 AA;
```

```
  Query Match             44.4%;  Score 20;  DB 20;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 SGAD 6
            ||||
Db        2 sgad 5


RESULT  26
Y42028
ID   Y42028 standard; Peptide; 7 AA.
XX
AC   Y42028;
XX
DT   09-DEC-1999  (first entry)
XX
DE   Rheumatoid arthritis diagnostic protein isoform peptide #179.
XX
KW   Human; rheumatoid arthritis; RA; diagnosis; RPI; RADF; detection;
KW   rheumatoid arthritis diagnostic feature; ERPI; synovial fluid;
KW   rheumatoid arthritis diagnostic protein isoform; screening;
KW   expression reference protein isoform; prognosis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 6
FT                   /label= Ile, Leu
XX
PN   WO9947925-A2.
XX
PD   23-SEP-1999.
XX
PF   15-MAR-1999;   99WO-GB00763.
XX
PR   13-MAR-1998;   98GB-0005477.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Parekh RB,  Patel TP,  Townsend RR;
XX
DR   WPI; 1999-571871/48.
XX
PT   Diagnosis of human rheumatoid arthritis by two-dimensional
PT   electrophoresis -
XX
PS   Disclosure; Page 21; 157pp; English.
XX
CC   A method has been developed for the diagnosis of human rheumatoid
CC   arthritis (RA) using two-dimensional electrophoresis to generate a
CC   two-dimensional array of features. The method can be used for screening,
CC   diagnosis and prognosis of RA in a subject or for monitoring the effect
CC   of an anti-RA drug or therapy administered to a subject. The method .
CC   comprises: (a) analysing a sample of serum or plasma and optionally
CC   synovial fluid by two-dimensional electrophoresis, to generate a two-
CC   dimensional array of features; (b) identifying at least one chosen
CC   feature whose relative abundance correlates with the presence or absence
CC   of RA; and (c) comparing the abundance of each chosen feature in the
CC   sample with the abundance of that chosen feature in serum or plasma from
CC   one or more persons without RA, where the relative abundance of the
CC   chosen feature or features in the sample indicates the presence or
CC   absence of RA in the subject. The method can also be used in clinical
CC   studies for testing drugs for therapy of RA, for purification of RA-
CC   diagnostic protein isoforms (RPIs), and for production of antibodies to
CC   RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify
CC   compounds that promote or inhibit their activity, which are then used as
CC   RA drugs. Nucleic acid encoding RADFs can be used in gene therapy
CC   protocols. Y41844 to Y42100 represent RPI peptides, Y42101 to Y42103
CC   represent expression reference protein isoform peptides and Z25066 to
CC   Z25068 represent degenerate probes for RPIs, which are all used in
CC   the exemplification of the present invention.
XX
```

```
SQ   Sequence   7 AA;

  Query Match             44.4%;  Score 20;  DB 20;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 SGAD 6
            ||||
Db        2 sgad 5


RESULT  27
W70077
ID   W70077 standard; peptide; 9 AA.
XX
AC   W70077;
XX
DT   28-OCT-1998  (first entry)
XX
DE   Thermus thermophilus methionyl-tRNA synthetase fragment.
XX
KW   Mycobacterium tuberculosis; mycobacterial; methionyl-tRNA synthetase;
KW   enzyme; antibiotic; tuberculosis; Thermus thermophilus.
XX
OS   Thermus thermophilus.
XX
PN   US5798240-A.
XX
PD   25-AUG-1998.
XX
PF   11-JAN-1996;   96US-0584226.
XX
PR   13-SEP-1994;   94US-0305766.
PR   11-JAN-1996;   96US-0584226.
XX
PA   (CUBI-) CUBIST PHARM INC.
XX
PI   Kim S,  Lee SH,  Martinis SA,  Sassanfar M,  Schimmel PR;
XX
DR   WPI; 1998-480383/41.
XX
PT   Recombinant genes encoding mycobacterial amino acyl tRNA synthetases
PT   - useful for recombinant production of the enzyme for use in
PT   screening of antibiotics against Mycobacterium tuberculosis
XX
PS   Example 1; Columns 33-34; 32pp; English.
XX
CC   Sequences shown in W70077 to W70086 represent methionyl-tRNA (Met-tRNA)
CC   synthetase fragments from different bacterial species. These are used
CC   for designing degenerate primers (V43887 to V43890) used for isolating
CC   Met-tRNa sythetase gene fragment by PCR from M. tuberculosis and M.
CC   kansasii. The invention provides an expression vector comprising a
CC   nucleic acid encoding a mycobacterial aminoacyl-tRNAs (atRNA) (especially
CC   Met-tRNA) synthetase under control of transcriptional signals that can be
CC   used to transform suitable host cells. The nucleict acids and host cells
CC   are used for the recombinant production of mycobacterial amino acyl tRNA
CC   synthetases, especially of methionyl tRNA synthetase. The enzymes are
CC   used by the organism in protein synthesis, and as such, the recombinant
CC   enzyme can be used to identify candidate drugs for use as antibiotics
CC   towards mycobacteria, especially M. tuberculosis, which is responsible
CC   for tuberculosis. Antisense constructs of the nucleic acids can also be
CC   used in antisense inhibition of the synthetase gene. The recombinant
CC   enzyme allows quick assays in screening of antibiotics. Present testing
CC   protocols involve exposing whole mycobacteria to candidate drugs and
CC   seeing their effect. This is time consuming as the bacteria are generally
CC   slow growing. Use of the enzyme also prevents researchers from having to
CC   work with pathogenic strains.
XX
SQ   Sequence   9 AA;

  Query Match             44.4%;  Score 20;  DB 19;  Length 9;
```

```
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGAD 6
           :|:| |
Db       1 fltgtd 6


RESULT  28
Y55545
ID   Y55545 standard; peptide; 9 AA.
XX
AC   Y55545;
XX
DT   17-JAN-2000  (first entry)
XX
DE   HLA binding plu-1 peptide.
XX
KW   Human; cancer-associated polypeptide; plu-1; ovarian cancer; vaccine;
KW   breast cancer; immune response; cytotoxic T lymphocyte; imaging agent;
KW   therapeutic; plu-1 antigen; breast tumor; tumor antigen; epitope.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9949034-A1.
XX
PD   30-SEP-1999.
XX
PF   19-MAR-1999;   99WO-GB00866.
XX
PR   20-MAR-1998;   98GB-0005877.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Taylor-papadimitriou J;
XX
DR   WPI; 1999-591090/50.
XX
PT   New nucleic acid encoding the cancer-associated polypeptide plu-1, for
PT   diagnosis, treatment and prevention of cancer, especially of breast and
PT   ovary  -
XX
PS   Example 2; Fig 12; 173pp; English.
XX
CC   The invention relates to a human cancer-associated polypeptide plu-1. The
CC   plu-1 polypeptide can be recombinantly expressed by standard recombinant
CC   methodology. Detection of the plu-1 nucleic acid or the polypeptide is
CC   used for the following: (i) diagnosis (including imaging) and prognosis
CC   of, and determination of susceptibility to, cancer, specifically ovarian
CC   or breast cancer; and ii) treating cancer (by inducing an immune response
CC   against cancer cells, e.g. as a vaccine, or by antisense inhibition).
CC   Antigens derived from the polypeptide are used to generate activated
CC   cytotoxic T lymphocytes, or dendritic cells, for subsequent return to
CC   the patient for treatment of cancer. The polypeptide may also be used to
CC   identify inhibitors of plu-1 activity. Fragments of the polypeptide, and
CC   antibodies raised against plu-1, are useful as assay and imaging agents,
CC   also therapeutically (to induce an anti-idiotype response or where
CC   conjugated to cytotoxic agents). The plu-1 antigen is expressed more
CC   commonly in breast tumors than some known tumor antigens. Sequences
CC   Y55320-629 represent predicted peptides from the plu-1 polypetide which
CC   may bind to the human class I alleles B27, A2, A3 and A11.
XX
SQ   Sequence   9 AA;


  Query Match             44.4%;  Score 20;  DB 20;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GADI 7
             ||||
Db       1 gadi 4
```

```
RESULT  29
Y55587
ID   Y55587 standard; peptide; 9 AA.
XX
AC   Y55587;
XX
DT   17-JAN-2000  (first entry)
XX
DE   HLA binding plu-1 peptide.
XX
KW   Human; cancer-associated polypeptide; plu-1; ovarian cancer; vaccine;
KW   breast cancer; immune response; cytotoxic T lymphocyte; imaging agent;
KW   therapeutic; plu-1 antigen; breast tumor; tumor antigen; epitope.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9949034-A1.
XX
PD   30-SEP-1999.
XX
PF   19-MAR-1999;   99WO-GB00866.
XX
PR   20-MAR-1998;   98GB-0005877.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Taylor-papadimitriou J;
XX
DR   WPI; 1999-591090/50.
XX
PT   New nucleic acid encoding the cancer-associated polypeptide plu-1, for
PT   diagnosis, treatment and prevention of cancer, especially of breast and
PT   ovary  -
XX
PS   Example 2; Fig 12; 173pp; English.
XX
CC   The invention relates to a human cancer-associated polypeptide plu-1. The
CC   plu-1 polypeptide can be recombinantly expressed by standard recombinant
CC   methodology. Detection of the plu-1 nucleic acid or the polypeptide is
CC   used for the following: (i) diagnosis (including imaging) and prognosis
CC   of, and determination of susceptibility to, cancer, specifically ovarian
CC   or breast cancer; and ii) treating cancer (by inducing an immune response
CC   against cancer cells, e.g. as a vaccine, or by antisense inhibition).
CC   Antigens derived from the polypeptide are used to generate activated
CC   cytotoxic T lymphocytes, or dendritic cells, for subsequent return to
CC   the patient for treatment of cancer. The polypeptide may also be used to
CC   identify inhibitors of plu-1 activity. Fragments of the polypeptide, and
CC   antibodies raised against plu-1, are useful as assay and imaging agents,
CC   also therapeutically (to induce an anti-idiotype response or where
CC   conjugated to cytotoxic agents). The plu-1 antigen is expressed more
CC   commonly in breast tumors than some known tumor antigens. Sequences
CC   Y55320-629 represent predicted peptides from the plu-1 polypetide which
CC   may bind to the human class I alleles B27, A2, A3 and A11.
XX
SQ   Sequence   9 AA;


  Query Match             44.4%;  Score 20;  DB 20;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GADI 7
             ||||
Db       3 gadi 6


RESULT  30
Y47818
ID   Y47818 standard; Peptide; 9 AA.
XX
```

```
AC   Y47818;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2429.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 124; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;


  Query Match             44.4%;  Score 20;  DB 20;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSGADI 7
            |||  |:
Db        1 ylsegdm 7


RESULT  31
Y05407
ID   Y05407 standard; peptide; 9 AA.
XX
AC   Y05407;
```

```
XX
DT   02-JUL-1999  (first entry)
XX
DE   Human BAK BH3 domain.
XX
KW   BH3 domain; cell death agonist; bcl homology domain; BCL-2 family;
KW   apoptosis promoter; cancer cell; virus infected cell; inflammation;
KW   autoantibody producing cell; cancer; lymphoproliferative condition;
KW   arthritis; autoimmune disease; therapy.
XX
OS   Homo sapiens.
XX
PN   WO9916787-A1.
XX
PD   08-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19765.
XX
PR   07-OCT-1997;   97US-0946039.
PR   26-SEP-1997;   97US-0060133.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Korsmeyer SJ;
XX
DR   WPI; 1999-255058/21.
XX
PT   Bcl homology domain 3 polypeptide
XX
PS   Claim 2; Fig 1; 104pp; English.
XX
CC   This sequence represents a bcl homology domain 3 (BH3 domain) of the
CC   invention, derived from a proapoptotic member of the BCL-2 family. The
CC   BH3 polypeptide can be used in a method for promoting apoptosis in a
CC   target cell, especially where the cell is a cancer cell a virus infected
CC   cell or an autoantibody producing cell. The BH3 polypeptide can be used
CC   in therapeutic compositions for treating disease including cancer, other
CC   lymphoproliferative conditions, arthritis, inflammation, and autoimmune
CC   diseases, which may result from the down regulation of cell death
CC   regulation.
XX
SQ   Sequence   9 AA;


  Query Match             44.4%;  Score 20;  DB 20;  Length 9;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 GADIN 8
            | |||
Db        5 gddin 9


RESULT  32
Y70831
ID   Y70831 standard; peptide; 9 AA.
XX
AC   Y70831;
XX
DT   31-JUL-2000  (first entry)
XX
DE   BH3 domain of human BAK protein.
XX
KW   Human; truncated BAX protein; tBAX; BAX alpha; BCL-2 family; head trauma;
KW   neuron; anti-apoptotic; cerebroprotective; neuroprotective; neuroactive;
KW   apoptosis; treatment; neurodegenerative disease; peripheral nerve injury;
KW   spinal cord injury; stroke; pro-apoptotic; PCD; programmed cell death;
KW   BH3 domain; BCL-2 homology domain; BAK protein.
XX
OS   Homo sapiens.
XX
PN   WO200023083-A1.
XX
```

```
PD   27-APR-2000.
XX
PF   22-OCT-1999;   99WO-US24747.
XX
PR   22-OCT-1998;   98US-0177315.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Johnson EM,  Easton R;
XX
DR   WPI; 2000-339513/29.
XX
PT   Truncated BAX polypeptides useful for preventing apoptosis of neurons
PT   for the treatment of nervous system disorders -
XX
PS   Disclosure; Fig 3; 43pp; English.
XX
CC   The patent discloses specific truncated BAX (tBAX) proteins
CC   which inhibit neuronal apoptosis induced by trophic factor deprivation.
CC   The anti-apoptotic tBAX proteins include tBAX70, tBAX78
CC   and their mutants. These proteins contain the N-terminal region and at
CC   least a portion of the BH3 domain of BAX alpha and lack the BH1, BH2 and
CC   C-terminal transmembrane domains. The tBAX protein lacking only the
CC   transmembrane domain has been shown to have anti-apoptotic activity.
CC   The tBAX proteins are used to treat diseases associated with neuronal
CC   apoptosis, e.g. neurodegenerative diseases, peripheral nerve injury,
CC   spinal cord injury, head trauma and stroke. tBAX proteins
CC   truncating at amino acid 68 of the BH3 domain of BAX alpha also have
CC   anti-apoptotic activity because Asp at position 68 is shown to be
CC   important for BAX activity and is conserved in all BCL-2 family members.
CC   The present sequence is a BH3 domain of human BAK protein, a
CC   pro-apoptotic protein belonging to BCL-2 family that is involved in
CC   regulation of neuronal programmed cell death. The present sequence
CC   contains the conserved Asp residue.
XX
SQ   Sequence   9 AA;


  Query Match             44.4%;  Score 20;  DB 21;  Length 9;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GADIN 8
           | |||
Db       5 gddin 9

RESULT  33
Y54210
ID   Y54210 standard; peptide; 9 AA.
XX
AC   Y54210;
XX
DT   06-APR-2000  (first entry)
XX
DE   HLA binding peptide 1331.05 derived from source p53.42A7.
XX
KW   Allele-specific binding motif; major histocompatibility complex; MHC;
KW   HLA; HLA-A2.1; cytotoxic T cell response; antigen; microbial infection;
KW   hepatitis; AIDS; malaria; condyloma acuminatum; cancer; prostate cancer;
KW   renal carcinoma; cervical carcinoma; lymphoma; tumour.
XX
OS   Unidentified.
XX
PN   WO9965522-A1.
XX
PD   23-DEC-1999.
XX
PF   17-JUN-1999;   99WO-US13789.
XX
PR   17-JUN-1998;   98US-0098584.
XX
PA   (EPIM-) EPIMMUNE INC.
```

```
XX
PI   Sette A,  Sidney J,  Southwood S;
XX
DR   WPI; 2000-106018/09.
XX
PT   Novel HLA binding immunogenic peptides used to induce T cell activation
PT   and to induce an immune response  -
XX
PS   Claim 1; Page 33; 42pp; English.
XX
CC   Peptides Y54171-Y54236 represent immunogenic peptides comprising an
CC   allele-specific binding motif for the major histocompatibility complex
CC   (MHC) molecule HLA, e.g. HLA-A2.1. The peptides have conserved residues
CC   at certain positions such as positions 2 and 9. Also, the peptides do not
CC   comprise negative binding residues at other positions, such as positions
CC   1, 3, 6 and/or 7 (peptides 9 amino acids long) and at poisons 1, 3, 4,
CC   5, 7, 8 and/or 9 (peptides'10 amino acids long). The peptides are used to
CC   induce a cytotoxic T cell response to a preselected antigen. The method
CC   comprises contacting cytotoxic T cells from a patient (optionally
CC   expressing a specific MHC class I allele) with the present peptides.
CC   The peptides are used to treat and prevent microbial infection (e.g. in
CC   viral hepatitis B and C, human papilomavirus (HPV) infection, AIDS,
CC   cytomegalovirus (CMV), malaria, and condyloma acuminatum) and cancer
CC   (e.g. prostate cancer, renal carcinoma, cervical carcinoma, lymphoma).
CC   Patients in the acute phase of infection can be treated with the
CC   peptides in conjunction with other treatments. The antigenic peptides
CC   may be used to elicit cytotoxic T lymphocytes (CTLs) ex vivo and in
CC   vivo. The resulting CTLs can be used to treat chronic infections (viral
CC   or bacterial) or tumours in patients that do not respond to conventional
CC   forms of therapy. The peptides may also be used to produce monoclonal
CC   antibodies, which are useful as potential diagnostic or therapeutic
CC   agents. The peptides may also be used as diagnostic reagents.
XX
SQ   Sequence   9 AA;


  Query Match             44.4%;  Score 20;  DB 21;  Length 9;
  Best Local Similarity   83.3%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LSGADI 7
           || |||
Db       4 lspadi 9

RESULT  34
W02264
ID   W02264 standard; peptide; 6 AA.
XX
AC   W02264;
XX
DT   22-OCT-1996  (first entry)
XX
DE   Gingivalis adhesion inhibitor comprising fimbrillin residues 240-245.
XX
KW   Fimbrillin gene; inhibition; adhesion; saliva; coated surface;
KW   prevention; periodontitis; teeth; gums; dentifrices; mouthwash;
KW   vaccine.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5536497-A.
XX
PD   16-JUL-1996.
XX
PF   21-DEC-1992;   92US-0994277.
XX
PR   21-DEC-1992;   92US-0994277.
XX
PA   (UYNY ) UNIV NEW YORK STATE RES FOUND.
XX
PI   Bedi GS,  Evans RT,  Genco RJ,  Sojar HT;
XX
```

```
DR   WPI; 1996-341445/34.
XX
PT   Peptide inhibitor of Porphyromonas gingivalis adhesion to saliva
PT   coated surface - useful for preventing periodontitis by application
PT   to the teeth and gums esp. in dentifrices, mouth:washes or topical
PT   formulations
XX
PS   Claim 1; Columns 17-18; 23pp; English.
XX
CC   The present peptide was prepd. by chemical synthesis, on the basis
CC   of an amino acid sequence deduced from the DNA sequence of the
CC   cloned P. gingivalis fimbrillin gene described in J. Bacteriol,
CC   170, 1658, 1988. The peptide inhibits the adhesion of P. gingivalis
CC   to saliva coated surfaces, and is therefore useful for preventing
CC   periodontitis by application to teeth and gums, esp. in
CC   dentifrices, mouthwashes or topical formulations, or by admin. as a
CC   vaccine. In an assay to determine the effect of the peptide on
CC   P. gingivalis binding to saliva coated hydroxapatite beads, the
CC   peptide resulted in a percentage binding inhibition of
CC   approx. 30 %.
XX
SQ   Sequence   6 AA;


  Query Match             42.2%;  Score 19;  DB 17;  Length 6;
  Best Local Similarity   60.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 SGADI 7
              :|||:
Db          1 ngadl 5


RESULT  35
W56879
ID   W56879 standard; peptide; 6 AA.
XX
AC   W56879;
XX
DT   28-JUL-1998  (first entry)
XX
DE   Enzyme inhibitor peptide SEQ ID NO:80.
XX
KW   Enzyme inhibitor; t-PA; u-PA; chymotrypsin; serine protease; active;
KW   latent; substrate subtraction phage display peptide library;
KW   identification; kinase; phosphatase; serpin.
XX
OS   Homo sapiens.
XX
PN   WO9747314-A1.
XX
PD   18-DEC-1997.
XX
PF   10-JUN-1997;   97WO-US09760.
XX
PR   10-JUN-1996;   96US-0019495.
XX
PA   (SCRI ) SCRIPPS RES INST.
XX
PI   Ke S,  Madison EL;
XX
DR   WPI; 1998-062746/06.
XX
PT   Substrate subtraction phage display peptide libraries - used to
PT   distinguish between active and latent forms of enzyme, e.g. serine
PT   protease
XX
PS   Claim 25; Page 67; 138pp; English.
XX
CC   The present sequence represents an enzyme inhibitor peptide used in
CC   the method of the invention to distinguish between t-PA and u-PA. The
CC   present invention describes a substrate subtraction library for the
CC   identification of peptide substrates selective between a first enzyme
CC   (E1) and a second enzyme (E2), comprising a collection different
CC   peptides, substantially lacking peptides that are effective substrates
CC   for E1. Also described are: (1) a method (M1) for identifying peptide
CC   substrates selective between a first enzyme (E1) and a second enzyme
CC   (E2); (2) a compound comprising the amino acid sequence of a peptide
CC   identified by M1; (3) a polypeptide for use as an enzyme inhibitor
CC   comprising one of 237 amino acid sequences (see W56801 to W56947, and
CC   W56949 to W57038); (4) a recombinant DNA vector comprising DNA (I)
CC   encoding a protease inhibitor including the sequence identified by the
CC   M1; (5) a prokaryotic or eukaryotic cell containing the vector of (4);
CC   (6) an antibody (Ab) immunoreactive with at least one of the peptides
CC   identified by M1; and (7) a diagnostic assay for distinguishing between
CC   active and latent forms of protease inhibitors, that uses (Ab). The
CC   library and method are used for distinguishing between active and latent
CC   forms of enzyme inhibitors, e.g. proteases, kinases and phosphatases.
CC   (Ab) are used for affinity purification of recombinant peptides and in
CC   the identification of naturally occurring protease inhibitors. Enzyme-
CC   inhibiting peptides identified can be used to treat a serpin deficiency
CC   or a disorder of serine proteases.
XX
SQ   Sequence   6 AA;

  Query Match             42.2%;  Score 19;  DB 19;  Length 6;
  Best Local Similarity   60.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 GADIN 8
              | |:|
Db          2 gvdmn 6


Search completed: December 16, 2000, 03:07:18
Job time: 8059 sec
```

```
                    GenCore version 4.5
           Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 02:00:56 ; Search time 113.2 Seconds
                                              (without alignments)
                                              7.424 Million cell updates/sec

Title:           US-09-529-121-2
Perfect score:   45
Sequence:        1 YLSGADLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters:      297973

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_14:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                %
Result          Query
  No.   Score  Match Length DB  ID                         Description
--------------------------------------------------------------------------
    1      35   77.8    519  2  Q47916                     Q47916 fibrobacter
    2      34   75.6    540  2  O31716                     O31716 bacillus su
    3      34   75.6    763  6  O62831                     O62831 bos taurus
    4      34   75.6    773  6  O62828                     O62828 bos taurus
    5      33   73.3    352 12   Q65258                     Q65258 african swi
    6      33   73.3    404 10  Q9SYB9                      Q9syb9 arabidopsis
    7      33   73.3    461  6  O97884                     O97884 equus cabal
    8      33   73.3    596  5  Q93896                     Q93896 caenorhabdi
    9      33   73.3    856 10  Q9SUN6                      Q9sun6 arabidopsis
   10      32   71.1    176  1  O58971                     O58971 pyrococcus
   11      32   71.1    350  5  Q20618                     Q20618 caenorhabdi
   12      32   71.1    377 10   Q9ZVI0                     Q9zvi0 arabidopsis
   13      32   71.1    391  1  O30217                     O30217 archaeoglob
   14      32   71.1    392  2  Q9RLU0                     Q9rlu0 listeria mo
   15      32   71.1    658  3  O74798                     O74798 schizosacch
   16      32   71.1    737  2  O51274                     O51274 borrelia bu
   17      32   71.1    763  4  Q16679                     Q16679 homo sapien
   18      32   71.1    836  5  Q9V9V6                     Q9v9v6 drosophila
   19      32   71.1    860  2  Q9WZL6                     Q9wzl6 thermotoga
   20      32   71.1    896  2  O50747                     O50747 borrelia bu
   21      32   71.1   1473  3  O14167                     O14167 schizosacch
   22      31   68.9    171  4  Q99521                     Q99521 homo sapien
   23      31   68.9    178 12   Q9QRQ7                     Q9qrq7 hepatitis b
   24      31   68.9    207 12   Q9QRQ8                     Q9qrq8 hepatitis b
   25      31   68.9    265  8  Q35783                     Q35783 sorghum bic
   26      31   68.9    295  2  Q9S578                     Q9s578 pseudomonas
   27      31   68.9    324  1  O57947                     O57947 pyrococcus
   28      31   68.9    363  5  O76806                     O76806 pisaster oc
   29      31   68.9    385  2  P94599                     P94599 bacillus th
   30      31   68.9    401  5  O02314                     O02314 caenorhabdi
   31      31   68.9    402  5  O02315                     O02315 caenorhabdi
   32      31   68.9    441  5  Q19407                     Q19407 caenorhabdi
   33      31   68.9    445  5  Q19730                     Q19730 caenorhabdi
   34      31   68.9    454  5  Q19729                     Q19729 caenorhabdi
   35      31   68.9    472  5  Q20406                     Q20406 caenorhabdi
   36      31   68.9    655  2  Q9RYM2                     Q9rym2 deinococcus
   37      31   68.9    663  4  Q9ULC6                     Q9ulc6 homo sapien
   38      31   68.9    705  2  O31988                     O31988 bacillus su
   39      31   68.9    705  9  O64034                     O64034 bacteriopha
   40      31   68.9    707  2  O30671                     O30671 bacillus su
   41      31   68.9    727 10   P93755                     P93755 arabidopsis
   42      31   68.9    742 12   Q83100                     Q83100 unidentifie
   43      31   68.9    880  3  Q06708                     Q06708 saccharomyc
   44      31   68.9   1006 10   Q9SA76                     Q9sa76 arabidopsis
   45      31   68.9   1170  3  Q9UVJ5                     Q9uvj5 botrytis ci

                              ALIGNMENTS


RESULT   1
Q47916
ID   Q47916     PRELIMINARY;      PRT;   519 AA.
AC   Q47916;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ENDOGLUCANASE CELG.
OS   Fibrobacter succinogenes (Bacteroides succinogenes).
OC   Bacteria; Fibrobacteria; Fibrobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S85;
RX   MEDLINE; 97017599.
RA   Iyo A.H., Forsberg C.W.;
RT   "Endoglucanase G from Fibrobacter succinogenes S85 belongs to a class
RT   of enzymes characterized by a basic C-terminal domain.";
RL   Can. J. Microbiol. 42:934-943(1996).
DR   EMBL; U33887; AAB38548.1; -.
DR   HSSP; P17901; 1EDG.
DR   INTERPRO; IPR001547; -.
DR   PFAM; PF00150; cellulase; 1.
SQ   SEQUENCE   519 AA;  56848 MW;  B06D2113B10FF27E CRC64;


  Query Match             77.8%;  Score 35;  DB 2;  Length 519;
  Best Local Similarity   87.5%;  Pred. No. 37;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSGADLNL 9
            |||:||||
Db      334 LSGSDLNL 341


RESULT   2
O31716
ID   O31716     PRELIMINARY;      PRT;   540 AA.
AC   O31716;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update).
DE   YKPA PROTEIN.
```

```
GN   YKPA.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE; 98044033.
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=I168;
RX   MEDLINE; 90368558.
RA   Hemila H., Palva A., Paulin L., Arvidson S., Palva I.;
RT   "Secretory S complex of Bacillus subtilis: sequence analysis and
RT   identity to pyruvate dehydrogenase.";
RL   J. Bacteriol. 172:5052-5063(1990).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=I168;
RX   MEDLINE; 97144523.
RA   Henriques A.O., Bryan E.M., Beall B.W., Moran C.P. Jr.;
RT   "cse15, cse60, and csk22 are new members of mother-cell-specific
RT   sporulation regulons in Bacillus subtilis.";
RL   J. Bacteriol. 179:389-398(1997).
DR   EMBL; Z99111; CAB13316.1; -.
DR   EMBL; AF012285; AAC24918.1; -.
DR   INTERPRO; IPR001617; -.
DR   PFAM; PF00005; ABC_tran; 2.
SQ   SEQUENCE   540 AA;  61055 MW;  F5588EDF1B948109 CRC64;

  Query Match             75.6%;  Score 34;  DB 2;  Length 540;
  Best Local Similarity   66.7%;  Pred. No. 63;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSGADLNL 9
            |  |:||||
Db      394 YFEGSDLNL 402

RESULT   3
O62831
ID   O62831     PRELIMINARY;      PRT;   763 AA.
AC   O62831;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   GASTRIN BINDING PROTEIN-LIKE PRECURSOR (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Lemaitre G., Cailleret K., Chevet E., Dahan S., Bergeron J.J.,
RA   Katinka M.D.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ222637; CAA10897.1; -.
DR   HSSP; P14604; 2DUB.
DR   INTERPRO; IPR001128; -.
DR   INTERPRO; IPR001753; -.
DR   INTERPRO; IPR002135; -.
DR   PFAM; PF00378; ECH; 1.
DR   PFAM; PF00725; 3HCDH; 1.
DR   PROSITE; PS00067; 3HCDH; 1.
DR   PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
DR   PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
KW   Signal.
FT   SIGNAL        1     36       POTENTIAL.
FT   CHAIN        37   >763       POTENTIAL.
FT   NON_TER     763    763
SQ   SEQUENCE   763 AA;  83291 MW;  81E72BA3222D7FF2 CRC64;

  Query Match             75.6%;  Score 34;  DB 6;  Length 763;
  Best Local Similarity   55.6%;  Pred. No. 95;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADLNL 9
            :::|||||:
Db       98 FIAGADLNM 106

RESULT   4
O62828
ID   O62828     PRELIMINARY;      PRT;   773 AA.
AC   O62828;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   FGF-2 BINDING PROTEIN PRECURSOR.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Chevet E., Lemaitre G., Cailleret K., Daham S., Bergeron J.J.,
RA   Katinka M.D.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ003123; CAA05878.1; -.
DR   HSSP; P14604; 2DUB.
DR   INTERPRO; IPR001128; -.
DR   INTERPRO; IPR001753; -.
DR   INTERPRO; IPR002135; -.
DR   PFAM; PF00378; ECH; 1.
DR   PFAM; PF00725; 3HCDH; 1.
DR   PROSITE; PS00067; 3HCDH; 1.
```

```
DR   PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
DR   PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
KW   Signal.
FT   SIGNAL        1     46       POTENTIAL.
FT   CHAIN        47    773       POTENTIAL.
SQ   SEQUENCE   773 AA;  84815 MW;  53B860715681445C CRC64;

  Query Match             75.6%;  Score 34;  DB 6;  Length 773;
  Best Local Similarity   55.6%;  Pred. No. 96;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADLNL 9
            :::|||||:
Db      108 FIAGADLNM 116

RESULT   5
Q65258
ID   Q65258      PRELIMINARY;      PRT;   352 AA.
AC   Q65258;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   ORF L3R.
OS   African swine fever virus (ASFV).
OC   Viruses; dsDNA viruses, no RNA stage;
OC   African swine fever-like viruses.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MALAWI LIL20 /1;
RX   MEDLINE; 94014996.
RA   Vydelingum S., Baylis S.A., Bristow C., Smith G.L., Dixon L.K.;
RT   "Duplicated genes within the variable right end of the genome of a
RT   pathogenic isolate of African swine fever virus.";
RL   J. Gen. Virol. 74:2125-2130(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MALAWI LIL20 /1;
RX   MEDLINE; 94292916.
RA   Dixon L.K., Twigg S.R.F., Baylis S.A., Vydelingum S., Bristow C.,
RA   Hammond J.M., Smith G.L.;
RT   "Nucleotide sequence of a 55 kbp region from the right end of the
RT   genome of a pathogenic African swine fever virus isolate (Malawi
RT   LIL20/1).";
RL   J. Gen. Virol. 7:1655-1684(1994).
DR   EMBL; X71982; CAA50855.1; -.
DR   INTERPRO; IPR002595; -.
DR   PFAM; PF01671; ASFV_360; 1.
DR   PRODOM; PD003462; -; 1.
SQ   SEQUENCE   352 AA;  40682 MW;  61561D08AE1C1599 CRC64;

  Query Match             73.3%;  Score 33;  DB 12;  Length 352;
  Best Local Similarity   75.0%;  Pred. No. 62;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGADLN 8
            || |||:|
Db      258 YLLGADIN 265

RESULT   6
Q9SYB9
ID   Q9SYB9      PRELIMINARY;      PRT;   404 AA.
AC   Q9SYB9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   F11M15.3 PROTEIN.
GN   F11M15.3.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC   Brassicaceae; Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA   Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
RA   Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
RA   Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
RA   Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006085; AAD30630.1; -.
DR   INTERPRO; IPR000719; -.
DR   INTERPRO; IPR002290; -.
DR   PFAM; PF00069; pkinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
SQ   SEQUENCE   404 AA;  45664 MW;  17BD70E18A25B063 CRC64;

  Query Match             73.3%;  Score 33;  DB 10;  Length 404;
  Best Local Similarity   66.7%;  Pred. No. 73;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSGADLNL 9
            | || |||:
Db      107 YCSGGDLNV 115

RESULT   7
O97884
ID   O97884      PRELIMINARY;      PRT;   461 AA.
AC   O97884;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   MELANOCYTE PROTEIN 17 PRECURSOR (FRAGMENT).
GN   PMEL17.
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SKIN;
RA   Rieder S., Checa-Cortes M.L., Joerg H., Stranzinger G.;
RT   "An equine sequence homologous to melanocyte protein 17 (PMEL17)
RT   mapped to chromosome 6q23.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF076780; AAC97108.1; -.
DR   INTERPRO; IPR000601; -.
FT   NON_TER       1      1
FT   NON_TER     461    461
SQ   SEQUENCE   461 AA;  49334 MW;  12752AF6C1EC373D CRC64;

  Query Match             73.3%;  Score 33;  DB 6;  Length 461;
  Best Local Similarity   75.0%;  Pred. No. 85;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADLN 8
            ||:||||:
Db       59 YLAGADLS 66

RESULT   8
Q93896
ID   Q93896      PRELIMINARY;      PRT;   596 AA.
AC   Q93896;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   M03B6.2 PROTEIN.
GN   M03B6.2.
```

```
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Matthews L.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z78545; CAB01766.1; -.
DR   INTERPRO; IPR002897; -.
DR   PFAM; PF01587; MCT; 1.
SQ   SEQUENCE   596 AA;  63884 MW;  4EEFB6CB56CBC7B6 CRC64;


  Query Match             73.3%;  Score 33;  DB 5;  Length 596;
  Best Local Similarity   87.5%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGADLN 8
            ||| ||||
Db      344 YLSKADLN 351


RESULT   9
Q9SUN6
ID   Q9SUN6      PRELIMINARY;      PRT;   856 AA.
AC   Q9SUN6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   PUTATIVE SERINE PROTEINASE.
GN   F9F13.80.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC   Brassicaceae; Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Pohl T., Weizenegger T., Bancroft I., Mewes H.W.,
RA   Mayer K.F.X., Lemcke K., Schueller C.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL080253; CAB45809.1; -.
DR   INTERPRO; IPR000209; -.
DR   INTERPRO; IPR002465; -.
DR   PFAM; PF00082; Peptidase_S8; 3.
DR   PRINTS; PR00723; SUBTILISIN.
DR   PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
DR   PROSITE; PS00340; RECEPTOR_CYTOKINES_2; UNKNOWN_1.
SQ   SEQUENCE   856 AA;  91814 MW;  0C9FD624F5DF81E2 CRC64;


  Query Match             73.3%;  Score 33;  DB 10;  Length 856;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+02;
```

```
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSGADLNL 9
            :||:||||
Db      751 ISGSDLNL 758


RESULT  10
O58971
ID   O58971      PRELIMINARY;      PRT;   176 AA.
AC   O58971;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   176AA LONG HYPOTHETICAL THERMONUCLEASE.
GN   PH1212.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE; 98344137.
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000005; BAA30312.1; -.
DR   HSSP; P00644; 1SNQ.
DR   INTERPRO; IPR000080; -.
DR   INTERPRO; IPR002071; -.
DR   PFAM; PF00565; SNase; 1.
DR   PROSITE; PS01123; TNASE_1; 1.
DR   PRODOM; PD002274; -; 1.
SQ   SEQUENCE   176 AA;  20010 MW;  CB802A467B17E29E CRC64;


  Query Match             71.1%;  Score 32;  DB 1;  Length 176;
  Best Local Similarity   62.5%;  Pred. No. 45;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGADLN 8
            ||:| |:|
Db      128 YLNGTDIN 135


RESULT  11
Q20618
ID   Q20618      PRELIMINARY;      PRT;   350 AA.
AC   Q20618;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   F49E12.5 PROTEIN.
GN   F49E12.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Thomas K.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
```

```
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z66520; CAA91390.1; -.
DR   INTERPRO; IPR000344; -.
DR   INTERPRO; IPR002106; -.
DR   PFAM; PF02117; Sra; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
SQ   SEQUENCE   350 AA;  40044 MW;  F92F03B3C7347B68 CRC64;


  Query Match             71.1%;  Score 32;  DB 5;  Length 350;
  Best Local Similarity   66.7%;  Pred. No. 1e+02;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
                |:| || ||
Db        286 YMSSADYNL 294


RESULT  12
Q9ZVI0
ID   Q9ZVI0      PRELIMINARY;      PRT;   377 AA.
AC   Q9ZVI0;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   T6A23.22 PROTEIN.
GN   T6A23.22.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC   Brassicaceae; Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
RA   Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.;
RT   "Arabidopsis thaliana chromosome II BAC T6A23 genomic sequence.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005499; AAC67360.1; -.
SQ   SEQUENCE   377 AA;  41905 MW;  0D066A25C9A02930 CRC64;


  Query Match             71.1%;  Score 32;  DB 10;  Length 377;
  Best Local Similarity   85.7%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADL 7
                |:|||||
Db        367 YVSGADL 373


RESULT  13
O30217
ID   O30217      PRELIMINARY;      PRT;   391 AA.
AC   O30217;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   3-KETOACYL-COA THIOLASE (ACAB-1).
GN   AF0018.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC   Archaeoglobus.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE; 98049343.
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
DR   EMBL; AE001105; AAB91208.1; -.
DR   TIGR; AF0018; -.
DR   INTERPRO; IPR002155; -.
DR   PFAM; PF00108; thiolase; 1.
DR   PROSITE; PS00737; THIOLASE_2; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE   391 AA;  43065 MW;  F26932FA3A20E8C0 CRC64;


  Query Match             71.1%;  Score 32;  DB 1;  Length 391;
  Best Local Similarity   66.7%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
                |:||||:| |
Db        355 YVSGAELGL 363


RESULT  14
Q9RLU0
ID   Q9RLU0      PRELIMINARY;      PRT;   392 AA.
AC   Q9RLU0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   HYPOTHETICAL 43.6 KDA PROTEIN.
OS   Listeria monocytogenes.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Listeria.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 7973;
RA   Morse R., O'Hanlon K., Virji M., Collins M.D.;
RT   "Isolation of rifampin-resistant mutants of Listeria monocytogenes and
RT   their characterization by rpoB gene sequencing, temperature
RT   sensitivity for growth, and interaction with an epithelial cell
RT   line.";
RL   J. Clin. Microbiol. 37:2913-2919(1999).
DR   EMBL; Y16468; CAB56705.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   392 AA;  43566 MW;  398B953F0B668D9F CRC64;


  Query Match             71.1%;  Score 32;  DB 2;  Length 392;
  Best Local Similarity   62.5%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLN 8
                |::|:|||
Db        221 YVTGSDLN 228


RESULT  15
O74798
ID   O74798      PRELIMINARY;      PRT;   658 AA.
AC   O74798;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
```

```
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HYPOTHETICAL 57.9 KDA PROTEIN.
GN   SPBC2D10.04.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC   Schizosaccharomycetaceae; Schizosaccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   Wood V., Rajandream M.A., Barrell B.G., Taylor K., Harris D.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL031788; CAA21162.1; -.
SQ   SEQUENCE   658 AA;  72731 MW;  40EB066BB970B4F7 CRC64;

  Query Match             71.1%;  Score 32;  DB 3;  Length 658;
  Best Local Similarity   75.0%;  Pred. No. 2.1e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGADLN 8
              ||:| |||
Db      159 YLAGFDLN 166


Search completed: December 16, 2000, 02:00:57
Job time: 5931 sec
```

```
                          GenCore version 4.5
                Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 01:51:13 ; Search time 91.14 Seconds
                                               (without alignments)
                                               6.267 Million cell updates/sec

Title:           US-09-529-121-2
Perfect score:   45
Sequence:        1 YLSGADLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters:      182106

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_65:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                               SUMMARIES
                     %
Result             Query
  No.    Score   Match Length DB  ID                                Description
--------------------------------------------------------------------------------
    1      40    88.9    702  2  A36319                            carcinoembryonic a
    2      34    75.6    506  2  C81704                            monooxygenase-rela
    3      34    75.6    540  2  E69861                            ABC transporter (A
    4      33    73.3    491  2  A49179                            melanoma antigen h
    5      33    73.3    596  2  T23685                            hypothetical prote
    6      33    73.3    856  2  T10585                            serine proteinase
    7      32    71.1    108  2  S37313                            hlyU protein - Vib
    8      32    71.1    167  2  S29515                            lktC protein - Pas
    9      32    71.1    167  2  A30169                            hypothetical 19.9K
   10      32    71.1    176  2  F71064                            micrococcal nuclea
   11      32    71.1    275  2  PN0511                            gastrin-binding pr
   12      32    71.1    350  2  T22450                            hypothetical prote
   13      32    71.1    391  2  B69252                            3-ketoacyl-CoA thi
   14      32    71.1    502  2  S50519                            hypothetical prote
   15      32    71.1    658  2  T40107                            hypothetical 57.9
   16      32    71.1    661  1  KFHU13                            coagulation factor
   17      32    71.1    668  2  A46013                            coagulation factor
   18      32    71.1    714  2  C65007                            hypothetical prote
   19      32    71.1    737  2  C70132                            hypothetical prote
   20      32    71.1    763  2  JC2108                            long-chain-fatty-a
   21      32    71.1    860  2  C72338                            hypothetical prote
   22      32    71.1    896  2  H70233                            hypothetical prote
   23      32    71.1   1473  2  T38791                            hypothetical prote
   24      32    71.1   1986  2  S28353                            probable polyketid
   25      31    68.9    184  1  S47020                            ribosomal protein
   26      31    68.9    265  2  T14645                            hypothetical prote
   27      31    68.9    295  2  T44541                            hypothetical prote
   28      31    68.9    311  2  E43680                            D'311 protein - Af
   29      31    68.9    324  2  F71243                            probable translati
   30      31    68.9    360  2  S53536                            MAP kinase (EC 2.7
   31      31    68.9    360  2  I49066                            cell division cycl
   32      31    68.9    360  2  S52419                            protein kinase (EC
   33      31    68.9    361  2  A54805                            protein kinase Mpk
   34      31    68.9    399  2  A33396                            beta-N-acetylgluco
   35      31    68.9    401  2  T24929                            hypothetical prote
   36      31    68.9    402  2  T24930                            hypothetical prote
   37      31    68.9    408  2  A45507                            ribulose-bisphosph
   38      31    68.9    441  2  T20840                            hypothetical prote
   39      31    68.9    445  2  T21262                            hypothetical prote
   40      31    68.9    454  2  T21261                            hypothetical prote
   41      31    68.9    472  2  T22188                            hypothetical prote
   42      31    68.9    655  2  G75582                            cell division prot
   43      31    68.9    705  2  T12784                            sublancin 168 lant
   44      31    68.9    763  1  A49681                            long-chain-fatty-a
   45      31    68.9    778  2  A23308                            transcription init


                               ALIGNMENTS


RESULT   1
A36319
carcinoembryonic antigen precursor - human
N;Alternate names: CEA; meconium antigen 100
C;Species: Homo sapiens (man)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 31-Jan-2000
C;Accession: A36319; A27773; A31037; A25845; S08106; S31737; A44476; I54224; I59098;
R;Schrewe, H.; Thompson, J.; Bona, M.; Hefta, L.J.F.; Maruya, A.; Hassauer, M.; Shive
Mol. Cell. Biol. 10, 2738-2748, 1990
A;Title: Cloning of the complete gene for carcinoembryonic antigen: analysis of its p
A;Reference number: A36319; MUID:90258861
A;Accession: A36319
A;Molecule type: DNA
A;Residues: 1-702 <SCH>
A;Cross-references: GB:M17303; NID:g178676; PIDN:AAB59513.1; PID:g178677
A;Note: the authors show the codons TTA for residue 641-Phe and CAG for residue 646-T
R;Beauchemin, N.; Benchimol, S.; Cournoyer, D.; Fuks, A.; Stanners, C.P.
Mol. Cell. Biol. 7, 3221-3230, 1987
A;Title: Isolation and characterization of full-length functional cDNA clones for hum
A;Reference number: A27773; MUID:88038876
A;Accession: A27773
A;Molecule type: mRNA
A;Residues: 1-702 <BEA>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
R;Barnett, T.; Goebel, S.J.; Nothdurft, M.A.; Elting, J.J.
Genomics 3, 59-66, 1988
A;Title: Carcinoembryonic antigen family: characterization of cDNAs coding for NCA an
A;Reference number: A31037; MUID:89122014
A;Accession: A31037
A;Molecule type: mRNA
A;Residues: 1-702 <BAR>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
A;Note: the authors translated the codon GTG for residue 130 as Leu
R;Oikawa, S.; Nakazato, H.; Kosaki, G.
Biochem. Biophys. Res. Commun. 142, 511-518, 1987
A;Title: Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA
A;Reference number: A25845; MUID:87128144
A;Accession: A25845
A;Molecule type: mRNA
A;Residues: 5-702 <OIK>
A;Cross-references: GB:M15042; NID:g180198; PIDN:AAA51963.1; PID:g180199
R;Oikawa, S.
submitted to the EMBL Data Library, September 1989
A;Reference number: S08106
A;Accession: S08106
A;Molecule type: mRNA
A;Residues: 5-319,321-702 <OI2>
A;Cross-references: EMBL:X16455; NID:g29854; PIDN:CAA34474.1; PID:g825638
R;Barnett, T.
submitted to the EMBL Data Library, September 1991
A;Description: Genomic DNA sequence upstream of the translational start of the carcin
A;Reference number: S31737
```

```
A;Accession: S31737
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-141 <BA2>
A;Cross-references: EMBL:X62151
R;Khan, W.N.; Fraengsmyr, L.; Teglund, S.; Israelsson, A.; Bremer, K.; Hammarstroem, S.
Genomics 14, 384-390, 1992
A;Title: Identification of three new genes and estimation of the size of the carcinoemb
A;Reference number: A44476; MUID:93052339
A;Accession: A44476
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 35-141 <KHA>
R;Willcocks, T.C.; Craig, I.W.
Genomics 8, 492-500, 1990
A;Title: Characterization of the genomic organization of human carcinoembryonic antigen
A;Reference number: I54224; MUID:91139118
A;Accession: I54224
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-37 <RES>
A;Cross-references: GB:M60964; NID:g180215; PIDN:AAA51964.1; PID:g180217
R;Zimmermann, W.; Ortlieb, B.; Friedrich, R.; von Kleist, S.
Proc. Natl. Acad. Sci. U.S.A. 84, 2960-2964, 1987
A;Title: Isolation and characterization of cDNA clones encoding the human carcinoembryo
A;Reference number: I59098; MUID:87204247
A;Accession: I59098
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 331-702 <RE2>
A;Cross-references: GB:M16234; NID:g180240; PIDN:AAA51972.1; PID:g180241
R;Siepen, D.; Paxton, R.J.; Neumaier, M.; Shively, J.E.; Wagener, C.
Biochem. Biophys. Res. Commun. 147, 212-218, 1987
A;Title: Carcinoembryonic antigen (CEA) and two crossreacting antigens of 165 KD and 10
A;Reference number: A26831; MUID:87326349
A;Accession: A26831
A;Molecule type: protein
A;Residues: 35-64 <SIE>
R;Thomas, P.; Toth, C.A.
Biochem. Biophys. Res. Commun. 170, 391-396, 1990
A;Title: Carcinoembryonic antigen binding to Kupffer cells is via a peptide located at
A;Reference number: A35490; MUID:90321257
A;Accession: A35490
A;Molecule type: protein
A;Residues: 'X',140-151,'X',153,'X',155-156 <THO>
A;Note: this is the amino terminal end of a fragment shown to mediate uptake by Kupffer
C;Comment: This heavily glycosylated membane protein of unknown function is a widely us
C;Comment: This protein may be processed at its C-terminus. It is anchored to the membr
C;Genetics:
A;Gene: GDB:CEA
A;Cross-references: GDB:119054; OMIM:114890
A;Map position: 19q13.2-19q13.2
A;Introns: 22/1; 142/1; 235/1; 320/1; 413/1; 498/1; 591/1; 676/1
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termi
C;Keywords: blocked carboxyl end; glycoprotein; lipoprotein; membrane protein; phosphat
F;1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-678/Product: carcinoembryonic antigen #status predicted <MAT>
F;160-217/Domain: immunoglobulin homology <IMM1>
F;252-301/Domain: immunoglobulin homology <IMM2>
F;338-395/Domain: immunoglobulin homology <IMM3>
F;516-573/Domain: immunoglobulin homology <IMM4>
F;608-657/Domain: immunoglobulin homology <IMM5>
F;679-702/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F;678/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Gly) (in mature for


  Query Match             88.9%;  Score 40;  DB 2;  Length 702;
  Best Local Similarity   88.9%;  Pred. No. 3.5;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
                |||||:|||
```

```
Db        605 YLSGANLNL 613

RESULT   2
C81704
monooxygenase-related protein TC0425 [imported] - Chlamydia muridarum (strain Nigg)
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C;Accession: C81704
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A;Reference number: A81500; MUID:20150255
A;Accession: C81704
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-506 <TET>
A;Cross-references: GB:AE002309; GB:AE002160; NID:g7190464; PIDN:AAF39281.1; PID:g719
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
A;Gene: TC0425


  Query Match             75.6%;  Score 34;  DB 2;  Length 506;
  Best Local Similarity   66.7%;  Pred. No. 41;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
                |||| :||:
Db        290 YLSGVNLNI 298


RESULT   3
E69861
ABC transporter (ATP-binding protein) homolog ykpA - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 17-Mar-2000
C;Accession: E69861
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033
A;Accession: E69861
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-540 <KUN>
A;Cross-references: GB:Z99111; GB:AL009126; NID:g2633699; PIDN:CAB13316.1; PID:e11850
A;Experimental source: strain 168
C;Genetics:
A;Gene: ykpA
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
C;Keywords: ATP; P-loop
F;17-228/Domain: ATP-binding cassette homology <ABC1>
F;34-41/Region: nucleotide-binding motif A (P-loop)
F;335-512/Domain: ATP-binding cassette homology <ABC2>


  Query Match             75.6%;  Score 34;  DB 2;  Length 540;
  Best Local Similarity   66.7%;  Pred. No. 44;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy          1 YLSGADLNL 9
              |  |:||||
Db        394 YFEGSDLNL 402


RESULT   4
A49179
melanoma antigen homolog rpe1 - bovine (fragment)
C;Species: Bos primigenius taurus (cattle)
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Nov-1996
C;Accession: A49179; I45861
R;Kim, R.Y.; Wistow, G.J.
Exp. Eye Res. 55, 657-662, 1992
A;Title: The cDNA RPE1 and monoclonal antibody HMB-50 define gene products preferentiall
A;Reference number: A49179; MUID:93122163
A;Accession: A49179
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-491 <KIM>
A;Experimental source: retinal pigment epithelium
A;Note: sequence extracted from NCBI backbone (NCBIN:122438, NCBIP:122439)
C;Genetics:
A;Gene: RPE1


  Query Match             73.3%;  Score 33;  DB 2;  Length 491;
  Best Local Similarity   75.0%;  Pred. No. 64;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLN 8
              ||:||||:
Db         93 YLAGADLS 100


RESULT   5
T23685
hypothetical protein M03B6.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T23685
R;Matthews, L.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19782
A;Accession: T23685
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-596 <WIL>
A;Cross-references: EMBL:Z78545; PIDN:CAB01766.1; GSPDB:GN00028; CESP:M03B6.2
A;Experimental source: clone M03B6
C;Genetics:
A;Gene: CESP:M03B6.2
A;Map position: X
A;Introns: 36/1; 94/1; 273/3; 320/1; 414/3; 533/1


  Query Match             73.3%;  Score 33;  DB 2;  Length 596;
  Best Local Similarity   87.5%;  Pred. No. 79;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADLN 8
              ||| ||||
Db        344 YLSKADLN 351


RESULT   6
T10585
serine proteinase homolog F9F13.80 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 24-Nov-1999
C;Accession: T10585
R;Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16991
A;Accession: T10585
A;Molecule type: DNA
A;Residues: 1-856 <BEV>
A;Cross-references: EMBL:AL080253; GSPDB:GN00062; ATSP:F9F13.80
A;Experimental source: cultivar Columbia; BAC clone F9F13
C;Genetics:
A;Gene: ATSP:F9F13.80
A;Map position: 4
A;Introns: 84/2; 137/3; 271/2; 303/2; 327/3; 422/1; 533/3; 624/1; 718/1
C;Superfamily: subtilisin-like proteinase ag12; subtilisin homology


  Query Match             73.3%;  Score 33;  DB 2;  Length 856;
  Best Local Similarity   75.0%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LSGADLNL 9
              :||:||||
Db        751 ISGSDLNL 758


RESULT   7
S37313
hlyU protein - Vibrio cholerae
C;Species: Vibrio cholerae
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 26-Aug-1999
C;Accession: S37313
R;Williams, S.G.; Attridge, S.R.; Manning, P.A.
Mol. Microbiol. 9, 751-760, 1993
A;Title: The transcriptional activator HlyU of Vibrio cholerae: nucleotide sequence a
A;Reference number: S37312; MUID:94049116
A;Accession: S37313
A;Molecule type: DNA
A;Residues: 1-108 <WIL>
A;Cross-references: EMBL:X66866; NID:g403330; PIDN:CAA47336.1; PID:g403332
C;Genetics:
A;Gene: hlyU
C;Superfamily: arsenical resistance operon repressor


  Query Match             71.1%;  Score 32;  DB 2;  Length 108;
  Best Local Similarity   66.7%;  Pred. No. 20;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              || || :||
Db          3 YLKGAPMNL 11


RESULT   8
S29515
lktC protein - Pasteurella haemolytica
C;Species: Pasteurella haemolytica
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C;Accession: S29515
R;Lo, R.Y.C.; Strathdee, C.A.; Shewen, P.E.
Infect. Immun. 55, 1987-1996, 1987
A;Title: Nucleotide sequence of the leukotoxin genes of Pasteurella haemolytica A1.
A;Reference number: S29515; MUID:87306837
A;Accession: S29515
A;Molecule type: DNA
A;Residues: 1-167 <LOR>
A;Cross-references: EMBL:M20730; NID:g150492; PIDN:AAA25528.1; PID:g150493
C;Genetics:
A;Gene: lktC
C;Superfamily: hemolysin C


  Query Match             71.1%;  Score 32;  DB 2;  Length 167;
  Best Local Similarity   77.8%;  Pred. No. 32;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy          1 YLSGADLNL 9
              | | |||||
Db         55 YCSWADLNL 63


RESULT    9
A30169
hypothetical 19.9K protein (lktA 5' region) - Pasteurella haemolytica
C;Species: Pasteurella haemolytica
C;Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 16-Jul-1999
C;Accession: A30169; C35254
R;Highlander, S.K.; Chidambaram, M.; Engler, M.J.; Weinstock, G.M.
DNA 8, 15-28, 1989
A;Title: DNA sequence of the Pasteurella haemolytica leukotoxin gene cluster.
A;Reference number: A30169; MUID:89210283
A;Accession: A30169
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-167 <HIG>
A;Cross-references: GB:M24197; GB:M34943; GB:M34944; NID:g150511; PIDN:AAA25542.1; PID:g
R;Highlander, S.K.; Engler, M.J.; Weinstock, G.M.
J. Bacteriol. 172, 2343-2350, 1990
A;Title: Secretion and expression of the Pasteurella haemolytica leukotoxin.
A;Reference number: A35254; MUID:90236888
A;Accession: C35254
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-20 <HI2>
A;Cross-references: GB:M24197; GB:M34943; GB:M34944
C;Superfamily: hemolysin C


  Query Match             71.1%;  Score 32;  DB 2;  Length 167;
  Best Local Similarity   77.8%;  Pred. No. 32;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              | | |||||
Db         55 YCSWADLNL 63


RESULT   10
F71064
micrococcal nuclease (EC 3.1.31.1) PH1212 precursor - Pyrococcus horikoshii
N;Alternate names: thermonuclease homolog
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C;Accession: F71064
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137
A;Accession: F71064
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-176 <KAW>
A;Cross-references: GB:AP000005; NID:g3236132; PIDN:BAA30312.1; PID:g3257629
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1212
C;Superfamily: micrococcal nuclease
C;Keywords: hydrolase
F;1-27/Domain: signal sequence #status predicted <SIG>


  Query Match             71.1%;  Score 32;  DB 2;  Length 176;
  Best Local Similarity   62.5%;  Pred. No. 34;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADLN 8
              ||:| |:|
```

```
Db        128 YLNGTDIN 135


RESULT  11
PN0511
gastrin-binding protein precursor - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 17-Mar-1999
C;Accession: PN0511
R;Baldwin, G.S.; Casey, A.; Weinstock, J.
Biochem. Biophys. Res. Commun. 193, 560-564, 1993
A;Title: Partial structure of the gene encoding the 78KDa gastrin binding protein enc
A;Reference number: PN0511; MUID:93290643
A;Accession: PN0511
A;Molecule type: mRNA
A;Residues: 1-275 <BAL>
A;Note: complete nucleotide sequence not given
C;Genetics:
A;Introns: 23/1; 37/1; 60/3; 105/2; 151/3; 191/3; 226/1
C;Superfamily: enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; 3-hydroxyacyl-CoA
F;1-36/Domain: signal sequence #status predicted <SIG>
F;37-275/Product: gastrin-binding protein #status predicted <MAT>
F;62-218/Domain: enoyl-CoA hydratase homology <ECH>


  Query Match             71.1%;  Score 32;  DB 2;  Length 275;
  Best Local Similarity   44.4%;  Pred. No. 55;
  Matches    4;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              :::|||:|:
Db         98 FIAGADINM 106


RESULT  12
T22450
hypothetical protein F49E12.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T22450
R;Thomas, K.
submitted to the EMBL Data Library, October 1995
A;Reference number: Z19565
A;Accession: T22450
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-350 <WIL>
A;Cross-references: EMBL:Z66520; PIDN:CAA91390.1; GSPDB:GN00020; CESP:F49E12.5
A;Experimental source: clone F49E12
C;Genetics:
A;Gene: CESP:F49E12.5
A;Map position: 2
A;Introns: 97/3; 171/3; 238/2; 300/3


  Query Match             71.1%;  Score 32;  DB 2;  Length 350;
  Best Local Similarity   66.7%;  Pred. No. 71;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              |:| || ||
Db        286 YMSSADYNL 294


RESULT  13
B69252
3-ketoacyl-CoA thiolase (acaB-1) homolog - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C;Accession: B69252
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, U.F.; McDonald, L.
```

```
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaec
A;Reference number: A69250; MUID:98049343
A;Accession: B69252
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-391 <KLE>
A;Cross-references: GB:AE001105; GB:AE000782; NID:g2689428; PIDN:AAB91208.1; PID:g265063
C;Superfamily: probable 3-ketoacyl-CoA thiolase MTH1177


  Query Match             71.1%;  Score 32;  DB 2;  Length 391;
  Best Local Similarity   66.7%;  Pred. No. 80;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
           |:|||:| |
Db     355 YVSGAELGL 363


RESULT  14
S50519
hypothetical protein YEL070w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein N3810; hypothetical protein YNR073c
C;Species: Saccharomyces cerevisiae
C;Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 20-Jun-2000
C;Accession: S50519; S63406; S63411; S54055; S63890
R;Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A;Description: The sequence of S. cerevisiae cosmids 9669, 8334, 8199, and lambda clone
A;Reference number: S50428
A;Accession: S50519
A;Molecule type: DNA
A;Residues: 1-502 <DIE>
A;Cross-references: EMBL:U18795; GSPDB:GN00005; MIPS:YEL070w; NID:g603241; PIDN:AAB65017
A;Genetics: C05
R;Andre, B.; Iraqui Houssaini, I.; Urrestarazu, L.A.; Vissers, S.
submitted to the Protein Sequence Database, April 1996
A;Reference number: S62920
A;Accession: S63406
A;Molecule type: DNA
A;Residues: 1-502 <AND>
A;Cross-references: EMBL:Z71688; GSPDB:GN00014; MIPS:YNR073c; NID:g1302609; PIDN:CAA9635
A;Experimental source: strain S288C
A;Genetics: C14
R;Levesque, H.; Lepingle, A.; Nicaud, J.M.; Gaillardin, C.
submitted to the Protein Sequence Database, April 1996
A;Reference number: S63411
A;Accession: S63411
A;Molecule type: DNA
A;Residues: 1-438 <LEV>
A;Cross-references: EMBL:Z71688; GSPDB:GN00014; MIPS:YNR073c; NID:g1302609; PIDN:CAA9635
A;Experimental source: strain S288C
A;Genetics: C14
R;Levesque, H.; Nicaud, J.M.; Lepingle, A.; Gaillardin, C.
submitted to the EMBL Data Library, May 1995
A;Description: Sequence of a 9.2kb telomeric fragment from the right arm of S. cerevisia
A;Reference number: S54055
A;Accession: S54055
A;Molecule type: DNA
A;Residues: 1-438 <LEW>
A;Cross-references: EMBL:X86790
A;Genetics: C14
R;Levesque, H.; Lepingle, A.; Nicaud, J.M.; Gaillardin, C.
Yeast 12, 289-295, 1996
A;Title: Sequencing of a 9.2 kb telomeric fragment from the right arm of Saccharomyces c
A;Reference number: S63890; MUID:97060021
A;Accession: S63890
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-438 <LEF>
```

```
A;Cross-references: EMBL:X86790
A;Genetics: C14
C;Genetics: <C05>
A;Gene: MIPS:YEL070w
A;Map position: 5L
C;Genetics: <C14>
A;Gene: MIPS:YNR073c
A;Map position: 14R
C;Superfamily: conserved hypothetical protein YEL070w


  Query Match             71.1%;  Score 32;  DB 2;  Length 502;
  Best Local Similarity   62.5%;  Pred. No. 1e+02;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGADLN 8
           ||:| |:|
Db     423 YLTGVDMN 430


RESULT  15
T40107
hypothetical 57.9 kd protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T40107
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Taylor, K.; Harris, D.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21906
A;Accession: T40107
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-658 <WOO>
A;Cross-references: EMBL:AL031788; PIDN:CAA21162.1; GSPDB:GN00067; SPDB:SPBC2D10.04
A;Experimental source: strain 972h-; cosmid c2D10
C;Genetics:
A;Gene: SPDB:SPBC2D10.04
A;Map position: 2


  Query Match             71.1%;  Score 32;  DB 2;  Length 658;
  Best Local Similarity   75.0%;  Pred. No. 1.4e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGADLN 8
           ||:| |||
Db     159 YLAGFDLN 166


Search completed: December 16, 2000, 01:51:14
Job time: 7635 sec
```

```
                    GenCore version 4.5
            Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 03:05:18 ; Search time 69.45 Seconds
                                               (without alignments)
                                               4.140 Million cell updates/sec

Title:          US-09-529-121-2
Perfect score:  45
Sequence:       1 YLSGADLNL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters:      87993

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      SwissProt_39:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                               SUMMARIES
                     %
Result             Query
  No.    Score   Match Length DB  ID                      Description
-------------------------------------------------------------------------
     1      40    88.9    702  1  CCEM_HUMAN              P06731 homo sapien
     2      34    75.6    455  1  PEX3_PICPA              Q92262 pichia past
     3      33    73.3    491  1  PM17_BOVIN              Q06154 bos taurus
     4      33    73.3   4655  1  LRP2_HUMAN              P98164 homo sapien
     5      32    71.1    108  1  HLYU_VIBCH              P52695 vibrio chol
     6      32    71.1    167  1  LKC1_PASHA              P16533 pasteurella
     7      32    71.1    167  1  LKC3_PASHA              P55120 pasteurella
     8      32    71.1    167  1  LKCB_PASHA              P55121 pasteurella
     9      32    71.1    502  1  YEI0_YEAST              P39941 saccharomyc
    10      32    71.1    572  1  GPC5_HUMAN              P78333 homo sapien
    11      32    71.1    661  1  F13B_HUMAN              P05160 homo sapien
    12      32    71.1    668  1  F13B_MOUSE              Q07968 mus musculu
    13      32    71.1    714  1  YFCX_ECOLI              P77399 escherichia
    14      32    71.1    763  1  ECHA_HUMAN              P40939 homo sapien
    15      32    71.1    763  1  ECHA_PIG                Q29554 sus scrofa
    16      32    71.1   1986  1  WA_EMENI                Q03149 emericella
    17      31    68.9    184  1  RS13_SULAC              P39470 sulfolobus
    18      31    68.9    311  1  V311_ASFB7              P23163 african swi
    19      31    68.9    360  1  MK11_MOUSE              O08666 mus musculu
    20      31    68.9    360  1  MK14_CANFA              O02812 canis famil
    21      31    68.9    360  1  MK14_HUMAN              Q16539 h mitogen-a
    22      31    68.9    360  1  MK14_MOUSE              P47811 mus musculu
    23      31    68.9    360  1  MK14_RAT                P70618 rattus norv
    24      31    68.9    361  1  MK14_CYPCA             Q90336 cyprinus ca
    25      31    68.9    361  1  MK14_XENLA             P47812 xenopus lae
    26      31    68.9    399  1  NALS_MOUSE              P15535 mus musculu
    27      31    68.9    408  1  RCA_CHLRE               P23489 chlamydomon
    28      31    68.9    763  1  ECHA_RAT                Q64428 rattus norv
    29      31    68.9    778  1  RAD3_YEAST              P06839 saccharomyc
    30      31    68.9    880  1  YL86_YEAST              Q06708 saccharomyc
    31      31    68.9   1006  1  BGAL_LACDE              P20043 lactobacill
    32      31    68.9   1096  1  KPC1_ASPNG              Q00078 aspergillus
    33      31    68.9   1139  1  KPC1_TRIRE              Q99014 trichoderma
```

```
    34      31    68.9   1142  1  KPC1_NEUCR              P87253 neurospora
    35      31    68.9   1174  1  KPC1_COCHE              O42632 cochliobolu
    36      31    68.9   2109  1  RRPL_VSVJH              P13615 vesicular s
    37      30    66.7    208  1  HBGF_MOUSE              Q06186 mus musculu
    38      30    66.7    299  1  PYRD_ARCFU              O29513 archaeoglob
    39      30    66.7    301  1  OTC_METTH               O27495 methanobact
    40      30    66.7    309  1  YF44_METTH              O27587 methanobact
    41      30    66.7    363  1  PGL1_COLLN              Q00446 colletotric
    42      30    66.7    408  1  YZ33_METJA              Q60290 methanococc
    43      30    66.7    414  1  PRS6_CAEEL              P46502 caenorhabdi
    44      30    66.7    418  1  PRS6_HUMAN              P43686 homo sapien
    45      30    66.7    418  1  PRS6_MOUSE              P54775 mus musculu

                               ALIGNMENTS

RESULT  1
CCEM_HUMAN
ID   CCEM_HUMAN     STANDARD;      PRT;   702 AA.
AC   P06731;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINOEMBRYONIC ANTIGEN PRECURSOR (CEA) (MECONIUM ANTIGEN 100)
DE   (CD66E ANTIGEN).
GN   CEA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90258861.
RA   Schrewe H., Thompson J., Bona M., Hefta L.J.F., Maruya A.,
RA   Hassauer M., Shively J.E., von Kleist S., Zimmermann W.;
RT   "Cloning of the complete gene for carcinoembryonic antigen: analysis
RT   of its promoter indicates a region conveying cell type-specific
RT   expression.";
RL   Mol. Cell. Biol. 10:2738-2748(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88038876.
RA   Beauchemin N., Benchimol S., Cournoyer D., Fuks A., Stanners C.P.;
RT   "Isolation and characterization of full-length functional cDNA clones
RT   for human carcinoembryonic antigen.";
RL   Mol. Cell. Biol. 7:3221-3230(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89122014.
RA   Barnett T., Goebel S.J., Nothdurft M.A., Elting J.J.;
RT   "Carcinoembryonic antigen family: characterization of cDNAs coding
RT   for NCA and CEA and suggestion of nonrandom sequence variation in
RT   their conserved loop-domains.";
RL   Genomics 3:59-66(1988).
RN   [4]
RP   SEQUENCE OF 5-702 FROM N.A.
RX   MEDLINE; 87128144.
RA   Oikawa S., Nakazato H., Kosaki G.;
RT   "Primary structure of human carcinoembryonic antigen (CEA) deduced
RT   from cDNA sequence.";
RL   Biochem. Biophys. Res. Commun. 142:511-518(1987).
RN   [5]
RP   SEQUENCE OF 331-702 FROM N.A.
RX   MEDLINE; 87204247.
RA   Zimmermann W., Ortlieb B., Friedrich R., von Kleist S.;
RT   "Isolation and characterization of cDNA clones encoding the human
RT   carcinoembryonic antigen reveal a highly conserved repeating
RT   structure.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:2960-2964(1987).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC   -!- TISSUE SPECIFICITY: FOUND IN ADENOCARCINOMAS OF ENDODERMALLY
CC       DERIVED DIGESTIVE SYSTEM EPITHELIUM AND FETAL COLON.
CC   -!- PTM: COMPLEX IMMUNOREACTIVE GLYCOPROTEIN WITH A MW OF 180 KDA
```

```
CC       COMPRISING 60% CARBOHYDRATE.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       7 IG-LIKE DOMAINS. BELONGS TO THE CARCINOEMBRYONIC ANTIGEN
CC       SUBFAMILY.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD66e entry;
CC       WWW="HTTP://WWW.NCBI.NLM.NIH.GOV/PROW/CD/CD66E.HTM".
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M17303; AAB59513.1; -.
DR   EMBL; M59262; AAA62835.1; ALT_SEQ.
DR   EMBL; M59255; AAA62835.1; JOINED.
DR   EMBL; M59256; AAA62835.1; JOINED.
DR   EMBL; M59257; AAA62835.1; JOINED.
DR   EMBL; M59258; AAA62835.1; JOINED.
DR   EMBL; M59259; AAA62835.1; JOINED.
DR   EMBL; M59260; AAA62835.1; JOINED.
DR   EMBL; M59261; AAA62835.1; JOINED.
DR   EMBL; M59709; -; NOT_ANNOTATED_CDS.
DR   EMBL; M59710; -; NOT_ANNOTATED_CDS.
DR   EMBL; M29540; AAA51967.1; -.
DR   EMBL; X16455; CAA34474.1; -.
DR   EMBL; M15042; AAA51963.1; -.
DR   EMBL; M16234; AAA51972.1; -.
DR   PIR; A36319; A36319.
DR   MIM; 114890; -.
DR   INTERPRO; IPR003006; -.
DR   PFAM; PF00047; ig; 7.
KW   Immunoglobulin domain; Glycoprotein; GPI-anchor; Membrane; Signal.
FT   SIGNAL        1     34
FT   CHAIN        35      ?       CARCINOEMBRYONIC ANTIGEN.
FT   PROPEP        ?    702       REMOVED IN MATURE FORM.
FT   DOMAIN       35    144       IG-LIKE DOMAIN 1.
FT   DOMAIN      146    237       IG-LIKE DOMAIN 2.
FT   DOMAIN      238    322       IG-LIKE DOMAIN 3.
FT   DOMAIN      324    415       IG-LIKE DOMAIN 4.
FT   DOMAIN      416    498       IG-LIKE DOMAIN 5.
FT   DOMAIN      502    593       IG-LIKE DOMAIN 6.
FT   DOMAIN      594    677       IG-LIKE DOMAIN 7.
FT   CARBOHYD    104    104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    115    115       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    152    152       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    182    182       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    197    197       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    204    204       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    208    208       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    246    246       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    256    256       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    274    274       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    288    288       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    292    292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    309    309       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    330    330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    351    351       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    360    360       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    375    375       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    432    432       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    466    466       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    480    480       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    508    508       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    529    529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    553    553       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    560    560       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    580    580       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    612    612       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    650    650       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    665    665       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CONFLICT    320    320       MISSING (IN REF. 4).
SQ   SEQUENCE   702 AA;  76795 MW;  6299AE26CDDBDB5C CRC64;

  Query Match             88.9%;  Score 40;  DB 1;  Length 702;
  Best Local Similarity   88.9%;  Pred. No. 1.2;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              |||||:|||
Db        605 YLSGANLNL 613


RESULT   2
PEX3_PICPA
ID   PEX3_PICPA     STANDARD;      PRT;   455 AA.
AC   Q92262;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PEROXISOMAL MEMBRANE PROTEIN PAS2 (PEROXIN-3).
GN   PEX3 OR PAS2.
OS   Pichia pastoris (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Pichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97115764.
RA   Subramani S.;
RT   "Protein translocation into peroxisomes.";
RL   J. Biol. Chem. 271:32483-32486(1996).
CC   -!- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z72390; CAA96530.1; -.
KW   Transmembrane; Peroxisome.
FT   DOMAIN        1     15       MATRIX (POTENTIAL).
FT   TRANSMEM     16     33       POTENTIAL.
FT   DOMAIN       34    455       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   455 AA;  51973 MW;  6853C5EA5C67EC34 CRC64;


  Query Match             75.6%;  Score 34;  DB 1;  Length 455;
  Best Local Similarity   66.7%;  Pred. No. 14;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGADLNL 9
              ||: ||:||
Db        353 YLNNADINL 361


RESULT   3
PM17_BOVIN
ID   PM17_BOVIN     STANDARD;      PRT;   491 AA.
AC   Q06154;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   MELANOCYTE PROTEIN PMEL 17 (RETINAL PIGMENT EPITHELIAL-SPECIFIC
DE   PROTEIN) (FRAGMENT).
GN   PMEL17 OR RPE1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=RETINA;
RX   MEDLINE; 93122163.
RA   Kim R.Y., Wistow G.J.;
RT   "The cDNA RPE1 and monoclonal antibody HMB-50 define gene products
RT   preferentially expressed in retinal pigment epithelium.";
RL   Exp. Eye Res. 55:657-662(1992).
CC   -!- FUNCTION: COULD BE A MELANOGENIC ENZYME (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC   -!- TISSUE SPECIFICITY: RETINAL PIGMENT EPITHELIUM.
CC   -!- SIMILARITY: CONTAINS 1 POLYCYSTIC KIDNEY DISEASE DOMAIN (PKD).
CC   -!- SIMILARITY: BELONGS TO THE PMEL-17/NMB FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M81193; AAA30419.1; -.
KW   Transmembrane; Glycoprotein; Melanin biosynthesis; Repeat.
FT   NON_TER       1      1
FT   DOMAIN       <1    423       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    424    444       POTENTIAL.
FT   DOMAIN      445    491       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       60    150       PKD.
FT   DOMAIN      148    256       8 X 13 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT      148    160       1.
FT   REPEAT      161    173       2.
FT   REPEAT      174    186       3.
FT   REPEAT      187    199       4.
FT   REPEAT      200    212       5.
FT   REPEAT      213    225       6.
FT   REPEAT      232    243       7.
FT   REPEAT      244    256       8.
FT   DOMAIN      304    394       CYS-RICH.
FT   CARBOHYD    269    269       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    396    396       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   491 AA;  51669 MW;  2BBFE5DFBD397D6D CRC64;


  Query Match             73.3%;  Score 33;  DB 1;  Length 491;
  Best Local Similarity   75.0%;  Pred. No. 24;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADLN 8
            ||:||||:
Db       93 YLAGADLS 100


RESULT   4
LRP2_HUMAN
ID   LRP2_HUMAN     STANDARD;      PRT;  4655 AA.
AC   P98164; O00711; Q16215;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED PROTEIN 2 PRECURSOR (MEGALIN)
DE   (GLYCOPROTEIN 330).
GN   LRP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY;
RX   MEDLINE; 96305376.
RA   Hjaelm G., Murray E., Crumley G., Harazim W., Lundgren S., Onyango I.,
RA   Ek B., Larsson M., Juhlin C., Hellman P., Davis H., Aekerstroem G.,
RA   Rask L., Morse B.;
```

```
RT   "Cloning and sequencing of human gp330, a Ca(2+)-binding receptor
RT   with potential intracellular signaling properties.";
RL   Eur. J. Biochem. 239:132-137(1996).
RN   [2]
RP   SEQUENCE OF 2705-4453 FROM N.A.
RC   TISSUE=KIDNEY;
RA   Knaak C., Argraves W.S.;
RL   Submitted (DEC-1993) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 3833-4453 FROM N.A.
RC   TISSUE=KIDNEY;
RX   MEDLINE; 95048397.
RA   Korenberg J.R., Argraves K.M., Chen X.N., Tran H.,
RA   Strickland D.K., Argraves W.S.;
RT   "Chromosomal localization of human genes for the LDL receptor family
RT   member glycoprotein 330 (LRP2) and its associated protein RAP
RT   (LRPAP1).";
RL   Genomics 22:88-93(1994).
RN   [4]
RP   SEQUENCE OF 4139-4406 FROM N.A.
RX   MEDLINE; 94244704.
RA   Lundgren S., Hjaelm G., Hellman P., Ek B., Juhlin C., Rastad J.,
RA   Klareskog L., Aakerstrom G., Rask L.;
RT   "A protein involved in calcium sensing of the human parathyroid and
RT   placental cytotrophoblast cells belongs to the LDL-receptor protein
RT   superfamily.";
RL   Exp. Cell Res. 212:344-350(1994).
RN   [5]
RP   FUNCTION.
RX   MEDLINE; 95286588.
RA   Kounnas M.Z., Loukinova E.B., Stefansson S., Harmony J.A.K.,
RA   Brewer B.H., Strickland D.K., Argraves W.S.;
RT   "Identification of glycoprotein 330 as an endocytic receptor for
RT   apolipoprotein J/clusterin.";
RL   J. Biol. Chem. 270:13070-13075(1995).
CC   -!- FUNCTION: BINDS SPECIFICALLY CLUSTERIN WITH HIGH AFFINITY, BUT
CC       ALSO LIGANDS IN COMMON WITH OTHER FAMILY MEMBERS: PLASMINOGEN,
CC       EXTRACELLULAR MATRIX COMPONENTS, PLASMINOGEN ACTIVATOR-PLASMINOGEN
CC       ACTIVATOR INHIBITOR TYPE I COMPLEX, APOLIPOPROTEIN E-ENRICHED
CC       BETA-VLDL, LIPOPROTEIN LIPASE, LACTOFERRIN AND CALCIUM.
CC   -!- FUNCTION: RECEPTOR-MEDIATED UPTAKE OF POLYBASIC DRUGS SUCH AS
CC       APROTININ, AMINOGLYCOSIDES AND POLYMYXIN B (BY SIMILARITY).
CC   -!- FUNCTION: MAY PARTICIPATE IN REGULATION OF PARATHYROID-HORMONE AND
CC       PARA-THYROID-HORMONE-RELATED PROTEIN RELEASE.
CC   -!- SUBUNIT: FORMS A MULTIMERIC COMPLEX TOGETHER WITH A RECEPTOR-
CC       ASSOCIATED PROTEIN (RAP).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: ABSORPTIVE EPITHELIA, INCLUDING RENAL
CC       PROXIMAL TUBULES.
CC   -!- SIMILARITY: CONTAINS 36 LDL-RECEPTOR CLASS A DOMAINS.
CC   -!- SIMILARITY: CONTAINS 37 LDL-RECEPTOR CLASS B DOMAINS.
CC   -!- SIMILARITY: CONTAINS 17 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 4 SH3-BINDING DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SH2-BINDING DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U33837; AAB41649.1; -.
DR   EMBL; U04441; AAB02882.1; -.
DR   EMBL; S73145; AAB30825.1; -.
DR   HSSP; P01130; 1AJJ.
DR   MIM; 600073; -.
DR   INTERPRO; IPR000033; -.
DR   INTERPRO; IPR000152; -.
DR   INTERPRO; IPR000561; -.
DR   INTERPRO; IPR001881; -.
DR   INTERPRO; IPR002172; -.
```

```
DR   PFAM; PF00008; EGF; 13.
DR   PFAM; PF00057; ldl_recept_a; 36.
DR   PFAM; PF00058; ldl_recept_b; 37.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   PROSITE; PS00010; ASX_HYDROXYL; 4.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 9.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PROSITE; PS01209; LDLRA_1; 31.
DR   PROSITE; PS50068; LDLRA_2; 36.
KW   Glycoprotein; Repeat; Endocytosis; Coated pits; Transmembrane;
KW   Receptor; EGF-like domain; Signal; Polymorphism.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   4655       LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED
FT                                PROTEIN 2.
FT   DOMAIN       26   4423       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   4424   4446       POTENTIAL.
FT   DOMAIN     4447   4655       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       26     64       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN       65    105       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      106    144       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      145    181       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN      182    219       LDL-RECEPTOR CLASS A 5.
FT   DOMAIN      220    258       LDL-RECEPTOR CLASS A 6.
FT   DOMAIN      264    308       LDL-RECEPTOR CLASS A 7.
FT   DOMAIN      309    347       EGF-LIKE 1.
FT   DOMAIN      348    386       EGF-LIKE 2.
FT   DOMAIN      436    477       LDL-RECEPTOR CLASS B 1.
FT   DOMAIN      479    520       LDL-RECEPTOR CLASS B 2.
FT   DOMAIN      522    567       LDL-RECEPTOR CLASS B 3.
FT   DOMAIN      569    612       LDL-RECEPTOR CLASS B 4.
FT   DOMAIN      613    653       LDL-RECEPTOR CLASS B 5.
FT   DOMAIN      659    705       EGF-LIKE 3.
FT   DOMAIN      753    794       LDL-RECEPTOR CLASS B 6.
FT   DOMAIN      796    836       LDL-RECEPTOR CLASS B 7.
FT   DOMAIN      838    880       LDL-RECEPTOR CLASS B 8.
FT   DOMAIN      882    924       LDL-RECEPTOR CLASS B 9.
FT   DOMAIN      970   1014       EGF-LIKE 4.
FT   DOMAIN     1024   1062       LDL-RECEPTOR CLASS A 8.
FT   DOMAIN     1065   1103       LDL-RECEPTOR CLASS A 9.
FT   DOMAIN     1107   1145       LDL-RECEPTOR CLASS A 10.
FT   DOMAIN     1147   1185       LDL-RECEPTOR CLASS A 11.
FT   DOMAIN     1186   1224       LDL-RECEPTOR CLASS A 12.
FT   DOMAIN     1228   1268       LDL-RECEPTOR CLASS A 13.
FT   DOMAIN     1269   1307       LDL-RECEPTOR CLASS A 14.
FT   DOMAIN     1310   1350       LDL-RECEPTOR CLASS A 15.
FT   DOMAIN     1349   1389       EGF-LIKE 5.
FT   DOMAIN     1390   1429       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1478   1519       LDL-RECEPTOR CLASS B 10.
FT   DOMAIN     1521   1562       LDL-RECEPTOR CLASS B 11.
FT   DOMAIN     1566   1608       LDL-RECEPTOR CLASS B 12.
FT   DOMAIN     1610   1653       LDL-RECEPTOR CLASS B 13.
FT   DOMAIN     1655   1695       LDL-RECEPTOR CLASS B 14.
FT   DOMAIN     1700   1741       EGF-LIKE 7.
FT   DOMAIN     1790   1831       LDL-RECEPTOR CLASS B 15.
FT   DOMAIN     1833   1881       LDL-RECEPTOR CLASS B 16.
FT   DOMAIN     1883   1929       LDL-RECEPTOR CLASS B 17.
FT   DOMAIN     1931   1971       LDL-RECEPTOR CLASS B 18.
FT   DOMAIN     1972   2012       LDL-RECEPTOR CLASS B 19.
FT   DOMAIN     2018   2059       EGF-LIKE 8.
FT   DOMAIN     2107   2155       LDL-RECEPTOR CLASS B 20.
FT   DOMAIN     2157   2200       LDL-RECEPTOR CLASS B 21.
FT   DOMAIN     2202   2244       LDL-RECEPTOR CLASS B 22.
FT   DOMAIN     2246   2288       LDL-RECEPTOR CLASS B 23.
FT   DOMAIN     2290   2331       LDL-RECEPTOR CLASS B 24.
FT   DOMAIN     2342   2383       EGF-LIKE 9.
FT   DOMAIN     2431   2476       LDL-RECEPTOR CLASS B 25.
FT   DOMAIN     2478   2517       LDL-RECEPTOR CLASS B 26.
FT   DOMAIN     2519   2561       LDL-RECEPTOR CLASS B 27.
FT   DOMAIN     2563   2603       LDL-RECEPTOR CLASS B 28.
FT   DOMAIN     2604   2646       LDL-RECEPTOR CLASS B 29.
FT   DOMAIN     2651   2693       EGF-LIKE 10.
FT   DOMAIN     2698   2738       LDL-RECEPTOR CLASS A 16.
FT   DOMAIN     2739   2777       LDL-RECEPTOR CLASS A 17.
FT   DOMAIN     2778   2819       LDL-RECEPTOR CLASS A 18.
FT   DOMAIN     2820   2861       LDL-RECEPTOR CLASS A 19.
FT   DOMAIN     2862   2901       LDL-RECEPTOR CLASS A 20.
FT   DOMAIN     2904   2945       LDL-RECEPTOR CLASS A 21.
FT   DOMAIN     2946   2990       LDL-RECEPTOR CLASS A 22.
FT   DOMAIN     2991   3029       LDL-RECEPTOR CLASS A 23.
FT   DOMAIN     3030   3070       LDL-RECEPTOR CLASS A 24.
FT   DOMAIN     3073   3110       LDL-RECEPTOR CLASS A 25.
FT   DOMAIN     3111   3151       EGF-LIKE 11.
FT   DOMAIN     3152   3192       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     3239   3280       LDL-RECEPTOR CLASS B 30.
FT   DOMAIN     3282   3331       LDL-RECEPTOR CLASS B 31.
FT   DOMAIN     3333   3375       LDL-RECEPTOR CLASS B 32.
FT   DOMAIN     3377   3418       LDL-RECEPTOR CLASS B 33.
FT   DOMAIN     3419   3459       LDL-RECEPTOR CLASS B 34.
FT   DOMAIN     3465   3509       EGF-LIKE 13.
FT   DOMAIN     3510   3550       LDL-RECEPTOR CLASS A 26.
FT   DOMAIN     3551   3591       LDL-RECEPTOR CLASS A 27.
FT   DOMAIN     3592   3632       LDL-RECEPTOR CLASS A 28.
FT   DOMAIN     3633   3673       LDL-RECEPTOR CLASS A 29.
FT   DOMAIN     3676   3716       LDL-RECEPTOR CLASS A 30.
FT   DOMAIN     3717   3756       LDL-RECEPTOR CLASS A 31.
FT   DOMAIN     3757   3795       LDL-RECEPTOR CLASS A 32.
FT   DOMAIN     3796   3834       LDL-RECEPTOR CLASS A 33.
FT   DOMAIN     3840   3880       LDL-RECEPTOR CLASS A 34.
FT   DOMAIN     3881   3922       LDL-RECEPTOR CLASS A 35.
FT   DOMAIN     3926   3964       LDL-RECEPTOR CLASS A 36.
FT   DOMAIN     3966   4006       EGF-LIKE 14.
FT   DOMAIN     4007   4048       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     4154   4195       LDL-RECEPTOR CLASS B 35.
FT   DOMAIN     4197   4239       LDL-RECEPTOR CLASS B 36.
FT   DOMAIN     4242   4282       LDL-RECEPTOR CLASS B 37.
FT   DOMAIN     4330   4368       EGF-LIKE 16.
FT   DOMAIN     4377   4411       EGF-LIKE 17.
FT   DOMAIN     4453   4459       SH3-BINDING (POTENTIAL).

  Query Match             73.3%;  Score 33;  DB 1;  Length 4655;
  Best Local Similarity   66.7%;  Pred. No. 2.8e+02;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
           : ||||||:
Db    4479 FRSGADLNM 4487


RESULT   5
HLYU_VIBCH
ID   HLYU_VIBCH     STANDARD;      PRT;   108 AA.
AC   P52695;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   TRANSCRIPTIONAL ACTIVATOR HLYU.
GN   HLYU.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLASSICAL 01 / Z17561;
RX   MEDLINE; 94049116.
RA   Williams S.G., Attridge S.R., Manning P.A.;
RT   "The transcriptional activator HlyU of Vibrio cholerae: nucleotide
RT   sequence and role in virulence gene expression.";
RL   Mol. Microbiol. 9:751-760(1993).
CC   -!- FUNCTION: UPREGULATES THE EXPRESSION OF THE HEMOLYSIN GENE, HLYA,
CC       AND MAY PROMOTE EXPRESSION OF OTHER VIRULENCE DETERMINANTS IN
CC       VIVO. IT MAY HAVE BOTH POSITIVE AND NEGATIVE REGULATOR ACTIVITIES.
CC   -!- SIMILARITY: BELONGS TO THE ARSR FAMILY OF TRANSCRIPTIONAL
CC       REGULATORS.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-9 IS THE INITIATOR.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X66866; CAA47336.1; -.
DR   HSSP; P30340; 1SMT.
DR   INTERPRO; IPR001845; -.
DR   PFAM; PF01022; HTH_5; 1.
DR   PRINTS; PR00778; HTHARSR.
KW   Transcription regulation; Activator; DNA-binding.
FT   DNA_BIND     47     66       H-T-H MOTIF (BY SIMILARITY).
SQ   SEQUENCE   108 AA;  12293 MW;  4FC1BE6DA960F217 CRC64;


  Query Match             71.1%;  Score 32;  DB 1;  Length 108;
  Best Local Similarity   66.7%;  Pred. No. 7.7;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
           || || :||
Db       3 YLKGAPMNL 11


RESULT   6
LKC1_PASHA
ID   LKC1_PASHA     STANDARD;      PRT;   167 AA.
AC   P16533;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   LEUKOTOXIN-ACTIVATING LYSINE-ACYLTRANSFERASE LKTC SEROTYPE A1
DE   (EC 2.3.1.-) (TOXIN-ACTIVATING PROTEIN C) (LEUKOTOXIN C).
GN   LKTC.
OS   Pasteurella haemolytica.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SEROTYPE A1;
RX   MEDLINE; 87306837.
RA   Lo R.Y.C., Strathdee C.A., Shewen P.E.;
RT   "Nucleotide sequence of the leukotoxin genes of Pasteurella
RT   haemolytica A1.";
RL   Infect. Immun. 55:1987-1996(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SEROTYPE A1 / PHL101;
RX   MEDLINE; 89210283.
RA   Highlander S.K., Chidambaram M., Engler M.J., Weinstock G.M.;
RT   "DNA sequence of the Pasteurella haemolytica leukotoxin gene
RT   cluster.";
RL   DNA 8:15-28(1989).
RN   [3]
RP   SEQUENCE OF 1-115 FROM N.A.
RC   STRAIN=SEROTYPE A1 / PHL101;
RX   MEDLINE; 93366458.
RA   Highlander S.K., Wickersham E.A.,.Garza O., Weinstock G.M.;
RT   "Expression of the Pasteurella haemolytica leukotoxin is inhibited by
RT   a locus that encodes an ATP-binding cassette homolog.";
RL   Infect. Immun. 61:3942-3951(1993).
RN   [4]
RP   ERRATUM.
RA   Highlander S.K., Wickersham E.A., Garza O., Weinstock G.M.;
RL   Infect. Immun. 61:5431-5431(1993).
RN   [5]
RP   SEQUENCE OF 1-20 FROM N.A.
RC   STRAIN=SEROTYPE A1 / PHL101;
RX   MEDLINE; 90236888.
RA   Highlander S.K., Engler M.J., Weinstock G.M.;
RT   "Secretion and expression of the Pasteurella haemolytica Leukotoxin.";
```

```
RL   J. Bacteriol. 172:2343-2350(1990).
CC   -!- FUNCTION: INVOLVED IN FATTY ACYLATION OF THE PROTOXIN (LKTA) AT
CC       TWO INTERNAL LYSINE RESIDUES, THEREBY CONVERTING IT TO THE ACTIVE
CC       TOXIN. THE ACYL DONOR IS ACP (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE RTX TOXIN ACYLTRANSFERASE FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between .the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M20730; AAA25528.1; -.
DR   EMBL; M24197; AAA25542.1; -.
DR   EMBL; M59210; AAA25537.1; -.
DR   PIR; S29515; S29515.
KW   Transferase; Acyltransferase; Hemolysis.
FT   ACT_SITE     22     22       PROBABLE.
FT   CONFLICT    157    157       A -> R (IN REF. 2).
SQ   SEQUENCE   167 AA;  19843 MW;  0D2ED2CBF2D7F1C5 CRC64;


  Query Match             71.1%;  Score 32;  DB 1;  Length 167;
  Best Local Similarity   77.8%;  Pred. No. 12;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
           | | |||||
Db      55 YCSWADLNL 63


RESULT   7
LKC3_PASHA
ID   LKC3_PASHA     STANDARD;      PRT;   167 AA.
AC   P55120;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   LEUKOTOXIN-ACTIVATING LYSINE-ACYLTRANSFERASE LKTC SEROTYPE T3
DE   (EC 2.3.1.-) (TOXIN-ACTIVATING PROTEIN C) (LEUKOTOXIN C).
GN   LKTC.
OS   Pasteurella haemolytica.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SEROTYPE T3;
RX   MEDLINE; 94041617.
RA   Burrows L.L., Olah-Winfield E., Lo R.Y.C.;
RT   "Molecular analysis of the leukotoxin determinants from Pasteurella
RT   haemolytica serotypes 1 to 16.";
RL   Infect. Immun. 61:5001-5007(1993).
CC   -!- FUNCTION: INVOLVED IN FATTY ACYLATION OF THE PROTOXIN (LKTA) AT
CC       TWO INTERNAL LYSINE RESIDUES, THEREBY CONVERTING IT TO THE ACTIVE
CC       TOXIN. THE ACYL DONOR IS ACP (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE RTX TOXIN ACYLTRANSFERASE FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U01216; AAB36690.1; -.
KW   Transferase; Acyltransferase; Hemolysis.
FT   ACT_SITE     22     22       PROBABLE.
SQ   SEQUENCE   167 AA;  19673 MW;  95B37E16C9D19C68 CRC64;
```

```
  Query Match             71.1%;  Score 32;  DB 1;  Length 167;
  Best Local Similarity   77.8%;  Pred. No. 12;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
           | | |||||
Db      55 YCSWADLNL 63

RESULT   8
LKCB_PASHA
ID   LKCB_PASHA     STANDARD;      PRT;   167 AA.
AC   P55121;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   LEUKOTOXIN-ACTIVATING LYSINE-ACYLTRANSFERASE LKTC SEROTYPE A11
DE   (EC 2.3.1.-) (TOXIN-ACTIVATING PROTEIN C) (LEUKOTOXIN C).
GN   LKTC.
OS   Pasteurella haemolytica.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SEROTYPE A11;
RX   MEDLINE; 94041617.
RA   Burrows L.L., Olah-Winfield E., Lo R.Y.C.;
RT   "Molecular analysis of the leukotoxin determinants from Pasteurella
RT   haemolytica serotypes 1 to 16.";
RL   Infect. Immun. 61:5001-5007(1993).
CC   -!- FUNCTION: INVOLVED IN FATTY ACYLATION OF THE PROTOXIN (LKTA) AT
CC       TWO INTERNAL LYSINE RESIDUES, THEREBY CONVERTING IT TO THE ACTIVE
CC       TOXIN. THE ACYL DONOR IS ACP (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE RTX TOXIN ACYLTRANSFERASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U01215; AAB36688.1; -.
KW   Transferase; Acyltransferase; Hemolysis.
FT   ACT_SITE     22     22       PROBABLE.
SQ   SEQUENCE   167 AA;  19753 MW;  0F4FA5D4E04CF795 CRC64;

  Query Match             71.1%;  Score 32;  DB 1;  Length 167;
  Best Local Similarity   77.8%;  Pred. No. 12;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
           | | |||||
Db      55 YCSWADLNL 63

RESULT   9
YEI0_YEAST
ID   YEI0_YEAST     STANDARD;      PRT;   502 AA.
AC   P39941;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL 56.5 KDA PROTEIN IN HXT8 5'REGION AND IN HXT17-COS10
DE   INTERGENIC REGION.
GN   (YEL070W) AND (YNR073C OR N3810).
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A. (YEL070W).
RC   STRAIN=S288C / AB972;
RA   Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA   Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA   Chung E., Duncan M., Guzman E., Hartzell G., Hunicke-Smith S.,
RA   Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA   Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA   Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA   Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A. (YNR073C).
RA   Andre B., Iraqui Houssaini I., Urrestarazu L.A., Vissers S.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-439 FROM N.A. (YNR073C).
RC   STRAIN=S288C;
RX   MEDLINE; 97060021.
RA   Levesque H., Nicaud J.-M., Lepingle A., Gaillardin C.;
RT   "Sequencing of a 9.2 kb telomeric fragment from the right arm of
RT   Saccharomyces cerevisiae chromosome XIV.";
RL   Yeast 12:289-295(1996).
CC   -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U18795; AAB65017.1; -.
DR   EMBL; Z71688; CAA96356.1; -.
DR   EMBL; Z71689; CAA96358.1; -.
DR   EMBL; X86790; CAA60486.1; -.
DR   SGD; S0000796; YEL070W.
DR   SGD; S0005356; YNR073C.
DR   INTERPRO; IPR000669; -.
DR   PFAM; PF01232; Mannitol_dh; 1.
DR   PRINTS; PR00084; MTLDHDRGNASE.
DR   PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW   Hypothetical protein; Oxidoreductase; NAD.
SQ   SEQUENCE   502 AA;  56470 MW;  A7E0CC01AADC1B2A CRC64;

  Query Match             71.1%;  Score 32;  DB 1;  Length 502;
  Best Local Similarity   62.5%;  Pred. No. 40;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGADLN 8
           ||:| |:|
Db     423 YLTGVDMN 430

RESULT  10
GPC5_HUMAN
ID   GPC5_HUMAN     STANDARD;      PRT;   572 AA.
AC   P78333; O60436;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLYPICAN-5 PRECURSOR.
GN   GPC5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 97224481.
```

```
RA   Veugelers M., Vermeesch J., Reekmans G., Steinfeld R., Marynen P.,
RA   David G.;
RT   "Characterization of glypican-5 and chromosomal localization of human
RT   GPC5, a new member of the glypican gene family.";
RL   Genomics 40:24-30(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 97470873.
RA   Saunders S., Paine-Saunders S., Lander A.D.;
RT   "Expression of the cell surface proteoglycan glypican-5 is
RT   developmentally regulated in kidney, limb, and brain.";
RL   Dev. Biol. 190:78-93(1997).
CC   -!- FUNCTION: HEPARAN SULFATE CELL SURFACE PROTEOGLYCAN (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY
CC       SIMILARITY).
CC   -!- TISSUE SPECIFICITY: IN ADULT, PRIMARILY EXPRESSED IN THE BRAIN.
CC       ALSO DETECTED IN FETAL BRAIN, LUNG, AND LIVER.
CC   -!- SIMILARITY: BELONGS TO THE GLYPICAN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U66033; AAC51118.1; -.
DR   EMBL; AF001462; AAC12261.1; -.
DR   MIM; 602446; -.
DR   INTERPRO; IPR001863; -.
DR   PFAM; PF01153; Glypican; 1.
DR   PROSITE; PS01207; GLYPICAN; 1.
KW   Proteoglycan; Heparan sulfate; Glycoprotein; Signal; GPI-anchor.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25      ?       GLYPICAN-5.
FT   PROPEP        ?    572       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       13     18       POLY-LEU.
FT   CARBOHYD    120    120       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    237    237       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    527    527       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    441    441       O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT   CARBOHYD    486    486       O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT   CARBOHYD    495    495       O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT   CARBOHYD    507    507       O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT   CARBOHYD    509    509       O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT   CONFLICT    205    205       G -> C (IN REF. 2).
FT   CONFLICT    245    245       S -> F (IN REF. 2).
SQ   SEQUENCE   572 AA;  63707 MW;  A17969FE0DD0D308 CRC64;


  Query Match             71.1%;  Score 32;  DB 1;  Length 572;
  Best Local Similarity   75.0%;  Pred. No. 46;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGADLN 8
          || |||:|
Db    151 YLFGADVN 158


RESULT 11
F13B_HUMAN
ID   F13B_HUMAN     STANDARD;      PRT;   661 AA.
AC   P05160;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   COAGULATION FACTOR XIII B CHAIN PRECURSOR (EC 2.3.2.13) (PROTEIN-
DE   GLUTAMINE GAMMA-GLUTAMYLTRANSFERASE B CHAIN) (TRANSGLUTAMINASE B
DE   CHAIN).
GN   F13B.
```

```
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91105054.
RA   Bottenus R.E., Ichinose A., Davie E.W.;
RT   "Nucleotide sequence of the gene for the b subunit of human factor
RT   XIII.";
RL   Biochemistry 29:11195-11209(1990).
RN   [2]
RP   SEQUENCE OF 2-661 FROM N.A.
RX   MEDLINE; 87026535.
RA   Ichinose A., McMullen B.A., Fujikawa K., Davie E.W.;
RT   "Amino acid sequence of the b subunit of human factor XIII, a protein
RT   composed of ten repetitive segments.";
RL   Biochemistry 25:4633-4638(1986).
RN   [3]
RP   REVISIONS.
RA   Ichinose A.;
RL   Submitted (FEB-1987) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1-20 FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 90251467.
RA   Grundmann U., Nerlich C., Rein T., Zettlmeissl G.;
RT   "Complete cDNA sequence encoding the B subunit of human factor XIII.";
RL   Nucleic Acids Res. 18:2817-2817(1990).
RN   [5]
RP   VARIANT PHE-450.
RX   MEDLINE; 93313189.
RA   Hashiguchi T., Saito M., Morishita E., Matsuda T., Ichinose A.;
RT   "Two genetic defects in a patient with complete deficiency of the b-
RT   subunit for coagulation factor XIII.";
RL   Blood 82:145-150(1993).
CC   -!- FUNCTION: THE B CHAIN OF FACTOR XIII IS NOT CATALYTICALLY ACTIVE,
CC       BUT IS THOUGHT TO STABILIZE THE A SUBUNITS AND REGULATE THE RATE
CC       OF TRANSGLUTAMINASE FORMATION BY THROMBIN.
CC   -!- SUBUNIT: TETRAMER OF TWO A CHAINS AND TWO B CHAINS.
CC   -!- DISEASE: A DEFICIENCY IN FA13 CAN RESULT IN A LIFELONG BLEEDING
CC       TENDENCY, DEFECTIVE WOUND HEALING, AND HABITUAL ABORTION.
CC   -!- SIMILARITY: CONTAINS 10 SUSHI (SCR) REPEATS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M64554; AAA51821.1; ALT_SEQ.
DR   EMBL; M14057; AAA88042.1; -.
DR   EMBL; X51823; CAA36123.1; -.
DR   PIR; A23830; A23830.
DR   PIR; A36397; A36397.
DR   PIR; S09980; S09980.
DR   HSSP; P08603; 1HFI.
DR   MIM; 134580; -.
DR   INTERPRO; IPR000436; -.
DR   PFAM; PF00084; sushi; 8.
KW   Transferase; Plasma; Blood coagulation; Repeat; Glycoprotein; Signal;
KW   Sushi; Disease mutation.
FT   SIGNAL        1     20
FT   CHAIN        21    661       COAGULATION FACTOR XIII B CHAIN.
FT   DOMAIN       24    647       10 X SUSHI (SCR) REPEATS.
FT   REPEAT       24     88       SUSHI 1.
FT   REPEAT       90    147       SUSHI 2.
FT   REPEAT      152    209       SUSHI 3.
FT   REPEAT      212    268       SUSHI 4.
FT   REPEAT      273    328       SUSHI 5.
FT   REPEAT      335    390       SUSHI 6.
FT   REPEAT      395    451       SUSHI 7.
```

```
FT   REPEAT      453    516       SUSHI 8.
FT   REPEAT      523    579       SUSHI 9.
FT   REPEAT      581    647       SUSHI 10.
FT   DISULFID     25     76       BY SIMILARITY.
FT   DISULFID     59     87       BY SIMILARITY.
FT   DISULFID     91    135       BY SIMILARITY.
FT   DISULFID    118    146       BY SIMILARITY.
FT   DISULFID    153    197       BY SIMILARITY.
FT   DISULFID    180    208       BY SIMILARITY.
FT   DISULFID    213    255       BY SIMILARITY.
FT   DISULFID    241    267       BY SIMILARITY.
FT   DISULFID    274    316       BY SIMILARITY.
FT   DISULFID    302    327       BY SIMILARITY.
FT   DISULFID    336    378       BY SIMILARITY.
FT   DISULFID    364    389       BY SIMILARITY.
FT   DISULFID    396    439       BY SIMILARITY.
FT   DISULFID    425    450       BY SIMILARITY.
FT   DISULFID    454    505       BY SIMILARITY.
FT   DISULFID    486    515       BY SIMILARITY.
FT   DISULFID    524    567       BY SIMILARITY.
FT   DISULFID    553    578       BY SIMILARITY.
FT   DISULFID    582    636       BY SIMILARITY.
FT   DISULFID    616    646       BY SIMILARITY.
FT   CARBOHYD    162    162       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    545    545       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE        617    619       CELL ATTACHMENT SITE.
FT   VARIANT     450    450       C -> F (IN F13B DEFICIENCY).
FT                                /FTId=VAR_007475.
SQ   SEQUENCE   661 AA;  75491 MW;  57A2FB46560857F2 CRC64;

  Query Match             71.1%;  Score 32;  DB 1;  Length 661;
  Best Local Similarity   85.7%;  Pred. No. 54;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADL 7
            ||||:||
Db      246 YLSGSDL 252


RESULT 12
F13B_MOUSE
ID   F13B_MOUSE     STANDARD;      PRT;   668 AA.
AC   Q07968;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   COAGULATION FACTOR XIII B CHAIN PRECURSOR (EC 2.3.2.13) (PROTEIN-
DE   GLUTAMINE GAMMA-GLUTAMYLTRANSFERASE B CHAIN) (TRANSGLUTAMINASE B
DE   CHAIN).
GN   F13B OR CF13B.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B10.D2/OSN; TISSUE=LIVER;
RX   MEDLINE; 93224141.
RA   Nonaka M., Matsuda Y., Shiroishi T., Moriwaki K., Nonaka M.,
RA   Natsuume-Sakai S.;
RT   "Molecular cloning of the b subunit of mouse coagulation factor XIII
RT   and assignment of the gene to chromosome 1: close evolutionary
RT   relationship to complement factor H.";
RL   Genomics 15:535-542(1993).
CC   -!- FUNCTION: THE B CHAIN OF FACTOR XIII IS NOT CATALYTICALLY ACTIVE,
CC       BUT IS THOUGHT TO STABILIZE THE A SUBUNITS AND REGULATE THE RATE
CC       OF TRANSGLUTAMINASE FORMATION BY THROMBIN.
CC   -!- SUBUNIT: TETRAMER OF TWO A CHAINS AND TWO B CHAINS.
CC   -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN LIVER AND KIDNEY.
CC   -!- SIMILARITY: CONTAINS 10 SUSHI (SCR) REPEATS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D10071; BAA00963.1; -.
DR   PIR; A46013; A46013.
DR   HSSP; P10998; 1VVC.
DR   MGD; MGI:88379; F13B.
DR   INTERPRO; IPR000436; -.
DR   PFAM; PF00084; sushi; 8.
KW   Transferase; Plasma; Blood coagulation; Repeat; Glycoprotein; Signal;
KW   Sushi.
FT   SIGNAL        1     20       BY SIMILARITY.
FT   CHAIN        21    668       COAGULATION FACTOR XIII B CHAIN.
FT   DOMAIN       24    647       10 X SUSHI (SCR) REPEATS.
FT   REPEAT       24     88       SUSHI 1.
FT   REPEAT       90    147       SUSHI 2.
FT   REPEAT      152    209       SUSHI 3.
FT   REPEAT      212    268       SUSHI 4.
FT   REPEAT      273    328       SUSHI 5.
FT   REPEAT      335    390       SUSHI 6.
FT   REPEAT      395    451       SUSHI 7.
FT   REPEAT      453    516       SUSHI 8.
FT   REPEAT      523    579       SUSHI 9.
FT   REPEAT      581    647       SUSHI 10.
FT   DISULFID     25     76       BY SIMILARITY.
FT   DISULFID     59     87       BY SIMILARITY.
FT   DISULFID     91    135       BY SIMILARITY.
FT   DISULFID    118    146       BY SIMILARITY.
FT   DISULFID    153    197       BY SIMILARITY.
FT   DISULFID    180    208       BY SIMILARITY.
FT   DISULFID    213    255       BY SIMILARITY.
FT   DISULFID    241    267       BY SIMILARITY.
FT   DISULFID    274    316       BY SIMILARITY.
FT   DISULFID    302    327       BY SIMILARITY.
FT   DISULFID    336    378       BY SIMILARITY.
FT   DISULFID    364    389       BY SIMILARITY.
FT   DISULFID    396    439       BY SIMILARITY.
FT   DISULFID    425    450       BY SIMILARITY.
FT   DISULFID    454    505       BY SIMILARITY.
FT   DISULFID    486    515       BY SIMILARITY.
FT   DISULFID    524    567       BY SIMILARITY.
FT   DISULFID    553    578       BY SIMILARITY.
FT   DISULFID    582    636       BY SIMILARITY.
FT   DISULFID    616    646       BY SIMILARITY.
FT   CARBOHYD    162    162       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    545    545       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   668 AA;  76078 MW;  80BC9E00A9E53FA6 CRC64;

  Query Match             71.1%;  Score 32;  DB 1;  Length 668;
  Best Local Similarity   85.7%;  Pred. No. 55;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADL 7
            ||||:||
Db      246 YLSGSDL 252


RESULT 13
YFCX_ECOLI
ID   YFCX_ECOLI     STANDARD;      PRT;   714 AA.
AC   P77399;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   PUTATIVE FATTY OXIDATION COMPLEX ALPHA SUBUNIT [INCLUDES: ENOYL-COA
DE   HYDRATASE (EC 4.2.1.17); 3-HYDROXYACYL-COA DEHYDROGENASE
DE   (EC 1.1.1.35); 3-HYDROXYBUTYRYL-COA EPIMERASE (EC 5.1.2.3)].
GN   YFCX.
OS   Escherichia coli.
```

```
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 97349980.
RA   Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA   Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA   Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA   Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,
RA   Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA   Yamagata S., Horiuchi T.;
RT   "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT   - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT   analysis of its sequence features.";
RL   DNA Res. 4:91-113(1997).
CC   -!- CATALYTIC ACTIVITY: (3S)-3-HYDROXYACYL-COA = TRANS-2(OR 3)-ENOYL-
CC       COA + H(2)O.
CC   -!- CATALYTIC ACTIVITY: L-3-HYDROXYACYL-COA + NAD(+) = 3-OXOACYL-COA +
CC       NADH.
CC   -!- CATALYTIC ACTIVITY: (S)-3-HYDROXYBUTANOYL-COA =
CC       (R)-3-HYDROXYBUTANOYL-COA.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE ENOYL-COA
CC       HYDRATASE/ISOMERASE FAMILY.
CC   -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE 3-
CC       HYDROXYACYL-COA DEHYDROGENASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE000322; AAC75401.1; -.
DR   EMBL; D90864; CAB22124.1; -.
DR   EMBL; D90865; CAB22130.1; -.
DR   HSSP; P14604; 2DUB.
DR   ECOGENE; EG14127; YFCX.
DR   INTERPRO; IPR001753; -.
DR   INTERPRO; IPR002135; -.
DR   PFAM; PF00725; 3HCDH; 1.
DR   PFAM; PF00378; ECH; 1.
DR   PROSITE; PS00067; 3HCDH; 1.
DR   PROSITE; PS00166; ENOYL_COA_HYDRATASE; FALSE_NEG.
KW   Hypothetical protein; Fatty acid metabolism; Multifunctional enzyme;
KW   Oxidoreductase; NAD; Lyase; Isomerase.
FT   ACT_SITE    118    118       ACTIVATES A WATER MOLECULE (BY
FT                                SIMILARITY).
FT   ACT_SITE    140    140       PROVIDES THE ALPHA-PROTON (BY
FT                                SIMILARITY).
SQ   SEQUENCE   714 AA;  77072 MW;  F4E0A75680039A0D CRC64;


  Query Match            71.1%;  Score 32;  DB 1;  Length 714;
  Best Local Similarity   44.4%;  Pred. No. 59;
  Matches    4;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADLNL 9
            :::|||:|:
Db       65 FIAGADINM 73
```

```
RESULT  14
ECHA_HUMAN
ID   ECHA_HUMAN     STANDARD;      PRT;   763 AA.
AC   P40939;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   MITOCHONDRIAL TRIFUNCTIONAL ENZYME ALPHA SUBUNIT PRECURSOR (TP-ALPHA)
DE   [INCLUDES: LONG-CHAIN ENOYL-COA HYDRATASE (EC 4.2.1.17); LONG CHAIN
DE   3-HYDROXYACYL-COA DEHYDROGENASE (EC 1.1.1.35)].
GN   HADHA OR HADH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94183263.
RA   Kamijo T., Aoyama T., Miyazaki J.I., Hashimoto T.;
RT   "Structural analysis of cDNAs for subunits of human mitochondrial
RT   fatty acid beta-oxidation trifunctional protein.";
RL   Biochem. Biophys. Res. Commun. 199:818-825(1994).
RN   [2]
RP   VARIANT AFLP GLN-510.
RX   MEDLINE; 95148633.
RA   Sims H.F., Brackett J.C., Powell C.K., Treem W.R., Hale D.E.,
RA   Bennett M.J., Gibson B., Shapiro S., Strauss A.W.;
RT   "The molecular basis of pediatric long chain 3-hydroxyacyl-CoA
RT   dehydrogenase deficiency associated with maternal acute fatty liver
RT   of pregnancy.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:841-845(1995).
CC   -!- FUNCTION: BIFUNCTIONAL SUBUNIT.
CC   -!- CATALYTIC ACTIVITY: (3S)-3-HYDROXYACYL-COA = TRANS-2(OR 3)-ENOYL-
CC       COA + H(2)O.
CC   -!- CATALYTIC ACTIVITY: L-3-HYDROXYACYL-COA + NAD(+) = 3-OXOACYL-COA +
CC       NADH.
CC   -!- PATHWAY: CATALYZES TWO STEPS OF THE FATTY ACID BETA-OXIDATION
CC       CYCLE.
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC   -!- DISEASE: DEFECTS IN HADHA ARE A CAUSE OF MATERNAL ACUTE FATTY
CC       LIVER OF PREGNANCY (AFLP), A SEVERE MATERNAL ILLNESS OCCURRING
CC       DURING PREGNANCIES WITH AFFECTED FETUSES. THIS DISEASE IS
CC       CHARACTERIZED BY SUDDEN UNEXPLAINED INFANT DEATH OR HYPOGLYCEMIA
CC       AND ABNORMAL LIVER ENZYMES (REYE-LIKE SYNDROME).
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE ENOYL-COA
CC       HYDRATASE/ISOMERASE FAMILY.
CC   -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE 3-
CC       HYDROXYACYL-COA DEHYDROGENASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D16480; BAA03941.1; -.
DR   PIR; JC2108; JC2108.
DR   HSSP; P14604; 2DUB.
DR   MIM; 600890; -.
DR   INTERPRO; IPR001753; -.
DR   INTERPRO; IPR002135; -.
DR   PFAM; PF00725; 3HCDH; 1.
DR   PFAM; PF00378; ECH; 1.
DR   PROSITE; PS00067; 3HCDH; 1.
DR   PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
KW   Fatty acid metabolism; Multifunctional enzyme; Oxidoreductase; NAD;
KW   Lyase; Mitochondrion; Transit peptide; Disease mutation.
FT   TRANSIT       1     36       MITOCHONDRION (POTENTIAL).
FT   CHAIN        37    763       MITOCHONDRIAL TRIFUNCTIONAL ENZYME ALPHA
FT                                SUBUNIT.
```

```
FT   ACT_SITE    151    151       ACTIVATES A WATER MOLECULE (BY
FT                                SIMILARITY).
FT   ACT_SITE    173    173       PROVIDES THE ALPHA-PROTON (BY
FT                                SIMILARITY).
FT   VARIANT     510    510       E -> Q (IN AFLP).
FT                                /FTId=VAR_002273.
SQ   SEQUENCE   763 AA;  82959 MW;  98D7881F7CB5949B CRC64;

  Query Match             71.1%;  Score 32;  DB 1;  Length 763;
  Best Local Similarity   44.4%;  Pred. No. 63;
  Matches    4;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
             :::|||:|:
Db      98 FIAGADINM 106


RESULT 15
ECHA_PIG
ID   ECHA_PIG       STANDARD;      PRT;   763 AA.
AC   Q29554;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   MITOCHONDRIAL TRIFUNCTIONAL ENZYME ALPHA SUBUNIT PRECURSOR (TP-ALPHA)
DE   (78 KDA GASTRIN-BINDING PROTEIN) [INCLUDES: LONG-CHAIN ENOYL-COA
DE   HYDRATASE (EC 4.2.1.17); LONG CHAIN 3-HYDROXYACYL-COA DEHYDROGENASE
DE   (EC 1.1.1.35)].
GN   HADHA OR LCHYD-HAD.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=STOMACH;
RX   MEDLINE; 94002212.
RA   Mantamadiotis T., Sobieszczuk P., Weinstock J., Baldwin G.S.;
RT   "Nucleotide sequence encoding a novel member of the
RT   hydratase/dehydrogenase family.";
RL   Biochim. Biophys. Acta 1170:211-215(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RX   MEDLINE; 94128080.
RA   Yang S.-Y., He X.-Y., Styles J., Luo M.J., Schulz H., Elzinga M.;
RT   "Primary structure of the large subunit of trifunctional beta-
RT   oxidation complex from pig heart mitochondria.";
RL   Biochem. Biophys. Res. Commun. 198:431-437(1994).
RN   [3]
RP   PARTIAL SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 93290643.
RA   Baldwin G.S., Casey A., Weinstock J.;
RT   "Partial structure of the gene encoding the 78 kDa gastrin binding
RT   protein excludes a close relationship with the peroxisomal
RT   trifunctional enzyme.";
RL   Biochem. Biophys. Res. Commun. 193:560-564(1993).
RN   [4]
RP   CHARACTERIZATION.
RC   TISSUE=HEART;
RX   MEDLINE; 95187623.
RA   Yang S.-Y.;
RT   "The large subunit of the pig heart mitochondrial membrane-bound beta-
RT   oxidation complex is a long-chain enoyl-CoA hydratase: 3-hydroxyacyl-
RT   CoA dehydrogenase bifunctional enzyme.";
RL   Comp. Biochem. Physiol. 109B:557-566(1994).
CC   -!- FUNCTION: BIFUNCTIONAL SUBUNIT; CANNOT USE CROTONYL-COA OR 3-
CC       HYDROXYBUTYRYL-COA AS SUBSTRATE.
CC   -!- CATALYTIC ACTIVITY: (3S)-3-HYDROXYACYL-COA = TRANS-2(OR 3)-ENOYL-
CC       COA + H(2)O.
CC   -!- CATALYTIC ACTIVITY: L-3-HYDROXYACYL-COA + NAD(+) = 3-OXOACYL-COA +
CC       NADH.
CC   -!- PATHWAY: CATALYZES TWO STEPS OF THE FATTY ACID BETA-OXIDATION
CC       CYCLE.
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE ENOYL-COA
CC       HYDRATASE/ISOMERASE FAMILY.
CC   -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE 3-
CC       HYDROXYACYL-COA DEHYDROGENASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L12581; AAA03733.1; -.
DR   EMBL; AF028609; AAB84118.1; -.
DR   HSSP; P14604; 2DUB.
DR   INTERPRO; IPR001753; -.
DR   INTERPRO; IPR002135; -.
DR   PFAM; PF00725; 3HCDH; 1.
DR   PFAM; PF00378; ECH; 1.
DR   PROSITE; PS00067; 3HCDH; 1.
DR   PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
KW   Fatty acid metabolism; Multifunctional enzyme; Oxidoreductase; NAD;
KW   Lyase; Mitochondrion; Transit peptide.
FT   TRANSIT       1     36       MITOCHONDRION (POTENTIAL).
FT   CHAIN        37    763       MITOCHONDRIAL TRIFUNCTIONAL ENZYME ALPHA
FT                                SUBUNIT.
FT   ACT_SITE    151    151       ACTIVATES A WATER MOLECULE (BY
FT                                SIMILARITY).
FT   ACT_SITE    173    173       PROVIDES THE ALPHA-PROTON (BY
FT                                SIMILARITY).
SQ   SEQUENCE   763 AA;  83106 MW;  DD0C6E7AE6B3D0F4 CRC64;

  Query Match             71.1%;  Score 32;  DB 1;  Length 763;
  Best Local Similarity   44.4%;  Pred. No. 63;
  Matches    4;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
             :::|||:|:
Db      98 FIAGADINM 106


Search completed: December 16, 2000, 03:05:19
Job time: 8039 sec
```

```
                    GenCore version 4.5
           Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 00:51:13 ; Search time 108.84 Seconds
                                              (without alignments)
                                              2.827 Million cell updates/sec

Title:           US-09-529-121-2
Perfect score:   45
Sequence:        1 YLSGADLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters:      268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_36:*
                1:  /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
                2:  /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
                3:  /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
                4:  /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
                5:  /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
                6:  /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
                7:  /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
                8:  /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
                9:  /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
               10:  /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
               11:  /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
               12:  /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
               13:  /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
               14:  /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
               15:  /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
               16:  /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
               17:  /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
               18:  /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
               19:  /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
               20:  /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
               21:  /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                 %
Result          Query
  No.   Score   Match Length DB  ID                         Description
 ------------------------------------------------------------------------------
    1      45   100.0      9 20  Y09526                     Carcinoembryonic a
    2      43    95.6      9 20  Y09527                     Carcinoembryonic a
    3      40    88.9      9 18  W39723                     Human carcina-embr
    4      40    88.9      9 19  W77134                     CEA synthetic pept
    5      40    88.9      9 19  W70045                     CEA derived HLA-A2
    6      40    88.9      9 20  Y47655                     Immunogenic peptid
    7      40    88.9      9 20  Y09525                     Carcinoembryonic a
    8      40    88.9     10 20  Y46555                     Immunogenic peptid
    9      40    88.9    107 20  W86133                     Protein sequence o
   10      40    88.9    178 10  P93499                     Sequence of carcin
   11      40    88.9    468 16  R77436                     BGP (1-314)/CEA (4
   12      40    88.9    493 16  R77435                     BGP (1-314)/CEA (4
```

```
    13      40   88.9    509  16  R77437                     BGP (1-314)/CEA (4
    14      40   88.9    511  16  R77438                     BGP (1-314)/CEA (4
    15      40   88.9    642  15  R60619                     Carcinoembryonic a
    16      40   88.9    663  17  R98519                     Immunogenic carcin
    17      40   88.9    698  9  P81229                     Carcinoembryonic a
    18      40   88.9    698  16  R65168                     Carcinoembryonic a
    19      40   88.9    698  18  W22844                     Human carcinoembry
    20      40   88.9    702  9  P81222                     Carcinoembryonic a
    21      40   88.9    702  10  P94014                     Carcinoembryonic c
    22      40   88.9    702  10  P93999                     Amino acid sequenc
    23      40   88.9    702  15  R54713                     Carcinoembryonic a
    24      40   88.9    702  17  W06872                     Carcinoembryonic a
    25      40   88.9    702  20  W83137                     CEA protein.  Homo
    26      40   88.9    734  17  W00182                     Carcinoembryonic a
    27      38   84.4      9  20  Y09528                     Carcinoembryonic a
    28      37   82.2      9  21  Y54173                     HLA binding peptid
    29      36   80.0      9  20  Y09529                     Carcinoembryonic a
    30      33   73.3      9  17  W00680                     Peptide comprising
    31      33   73.3    207  17  R97213                     Human calcium sens
    32      33   73.3    207  19  W26770                     Calcium sensor pro
    33      33   73.3    944  17  R97207                     Human calcium sens
    34      33   73.3    944  19  W43310                     Human placenta cal
    35      33   73.3   4655  17  R97208                     Human calcium sens
    36      33   73.3   4655  17  R97209                     Human placental ca
    37      33   73.3   4655  17  R97210                     Human kidney calci
    38      33   73.3   4655  17  R97211                     Human parathyroid
    39      33   73.3   4655  19  W43311                     Human calcium sens
    40      33   73.3   4655  19  W43312                     Human placental ca
    41      33   73.3   4655  19  W43314                     Human parathyroid
    42      32   71.1     81  15  R46083                     Enoyl CoA hydrase
    43      32   71.1    167  12  R14223                     Leukotoxin genes O
    44      32   71.1    167  14  R39632                     Leukotoxin protein
    45      32   71.1    572  20  Y28885                     Human glypican 5 p

                                 ALIGNMENTS

RESULT   1
Y09526
ID    Y09526 standard; peptide; 9 AA.
XX
AC    Y09526;
XX
DT    20-JUL-1999  (first entry)
XX
DE    Carcinoembryonic antigen peptide agonist SEQ ID NO:2.
XX
KW    Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW    immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW    bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW    adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS    Homo sapiens.
OS    Synthetic.
XX
PN    WO9919478-A1.
XX
PD    22-APR-1999.
XX
PF    22-SEP-1998;   98WO-US19794.
XX
PR    10-OCT-1997;   97US-0061589.
XX
PA    (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI    Barzaga E,  Schlom J,  Zaremba S;
XX
DR    WPI; 1999-326544/27.
XX
PT    Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS    Claim 5; Page 53; 72pp; English.
```

```
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             100.0%;  Score 45;  DB 20;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADLNL 9
            |||||||||
Db        1 ylsgadlnl 9

RESULT   2
Y09527
ID   Y09527 standard; peptide; 9 AA.
XX
AC   Y09527;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:3.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             95.6%;  Score 43;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADLNL 9
            ||||||:||
Db        1 ylsgadinl 9

RESULT   3
W39723
ID   W39723 standard; peptide; 9 AA.
XX
AC   W39723;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Human carcina-embryonic antigen (CEA) peptide (pos. 571-579).
XX
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
XX
OS   Homo sapiens.
XX
PN   WO9741440-A1.
XX
PD   06-NOV-1997.
XX
PF   28-APR-1997;   97WO-NL00229.
XX
PR   23-DEC-1996;   96EP-0203670.
PR   26-APR-1996;   96EP-0201145.
XX
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
PI   Kast WM,  Melief CJM,  Offringa R,  Toes REM,  Van Der Burg SH;
XX
DR   WPI; 1997-549891/50.
XX
PT   Method of selecting T cell peptide epitope(s) - by measuring the
PT   stability of HLA class I-peptide complexes on intact B cells
XX
PS   Example 3; Page 85; 109pp; English.
XX
CC   Peptides W39430-W39734 are used in a novel method for the selection of
CC   immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC   method involves the identification of peptide sequences capable of
CC   binding to an HLA (human leukocyte antigen) class I molecule and
CC   measuring the binding of this epitope peptide to the HLA class I peptide.
CC   The stability of binding of the peptide and MHC (major histocompatablilty
CC   complex) class I molecule is measured on intact human B cells carrying
CC   the MHC molecule at their cell surfaces. The method can be used to select
CC   peptide epitopes for generating vaccines against a disease associated
CC   with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC   especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC   immune responses. Peptide W39723 is derived from the human
CC   carcino-embryonic antigen (CEA) and has the ability to bind to the human
CC   MHC Class I allele HLA-A2.1.
XX
SQ   Sequence   9 AA;

  Query Match             88.9%;  Score 40;  DB 18;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
```

```
  Matches     8;  Conservative     1;  Mismatches     0;  Indels     0;  Gaps     0;

Qy       1 YLSGADLNL 9
           |||||:|||
Db       1 ylsganlnl 9


RESULT    4
W77134
ID   W77134 standard; peptide; 9 AA.
XX
AC   W77134;
XX
DT   16-NOV-1998  (first entry)
XX
DE   CEA synthetic peptide epitope 1.
XX
KW   Tyrosinase; tyrosinase cytotoxic lymphocyte response;
KW   cytotoxic T lymphocyte; cysteine-depleted; melanoma.
XX
OS   Synthetic.
XX
PN   WO9833810-A2.
XX
PD   06-AUG-1998.
XX
PF   29-JAN-1998;   98WO-US01592.
XX
PR   30-JAN-1997;   97US-0037781.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI   Engelhard VH,  Hunt DF,  Kittlesen D,  Slingluff CL;
XX
DR   WPI; 1998-437388/37.
XX
PT   Disease specific immunogen  - comprises disease specific cytotoxic T
PT   lymphocyte epitope used to elicit melanoma specific CTL response
XX
PS   Disclosure; Page 27; 93pp; English.
XX
CC   The peptide epitope W77119-W77138 were created for human tumour-specific
CC   cytotoxic T lymphocyte response.  These peptides are are cysteine-
CC   depleted mutants of a native disease-specific CTL epitope.  The cysteine-
CC   depleted CTL epitopes elicit a stronger or more specific CTL response
CC   than the native epitope.  The epitopes can be used in a disease-specific
CC   immunogen to protect a mammal against disease in particular melanomas.
CC   The peptides may also be used to screen a sample for the presence of
CC   an antigen with the same epitope, or with a different cross-reactive
CC   epitope.
XX
SQ   Sequence   9 AA;

  Query Match             88.9%;  Score 40;  DB 19;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches     8;  Conservative     1;  Mismatches     0;  Indels     0;  Gaps     0;

Qy       1 YLSGADLNL 9
           |||||:|||
Db       1 ylsganlnl 9


RESULT    5
W70045
ID   W70045 standard; peptide; 9 AA.
XX
AC   W70045;
XX
DT   22-OCT-1998  (first entry)
XX
DE   CEA derived HLA-A2.1 binding peptide 2 (residues 605-613).
XX
```

```
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW   antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW   viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW   fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9833888-A1.
XX
PD   06-AUG-1998.
XX
PF   30-JAN-1998;   98WO-US01959.
XX
PR   31-JAN-1997;   97US-0036696.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Celis E,  Sette A,  Sidney J,  Southwood S,  Tsai V;
XX
DR   WPI; 1998-437445/37.
XX
PT   Production of antigen-specific cytotoxic T cells - by incubating
PT   immunogenic peptide(s) from antigen that binds class I major
PT   histocompatibility complex molecules with pre-treated antigen
PT   presenting cells
XX
PS   Example 6; Page 75; 104pp; English.
XX
CC   Sequences shown in W70044 to W70052 represent peptides derived from
CC   carcinoembryonic antigen (CEA). The peptides can bind to a human
CC   leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method
CC   of invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC   vitro. The method comprises contacting immunogenic peptides from an
CC   antigen that binds class I major histocompatibility complex (MHC)
CC   molecules with antigen presenting cells (APCs) pretreated with
CC   pretreatment growth factors, and incubating the APCs with purified CD8
CC   cells in the presence of at least 2 incubation growth factors, thereby
CC   producing antigen-specific CTLs. A method for specifically killing
CC   target cells in a human patient is also provided which comprises
CC   obtaining a fluid sample containing CTLs from a patient, contacting the
CC   cytotoxic T cells with APCs pretreated with pre-treatment growth factors,
CC   where the APCs comprise class I MHC molecules. The pretreated APCs are
CC   incubated with the cytotoxic growth factors, thereby producing activated
CC   CTLs which are contacted with a carrier to form a composition. The
CC   composition can then be administered to the patient. The activated CTLs
CC   can be used for treating cancers, immune disorders, viral infections,
CC   AIDS, hepatitis, bacterial infection, fungal infection, malaria or
CC   tuberculosis.
XX
SQ   Sequence   9 AA;

  Query Match             88.9%;  Score 40;  DB 19;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches     8;  Conservative     1;  Mismatches     0;  Indels     0;  Gaps     0;

Qy       1 YLSGADLNL 9
           |||||:|||
Db       1 ylsganlnl 9


RESULT    6
Y47655
ID   Y47655 standard; Peptide; 9 AA.
XX
AC   Y47655;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2266.
XX
```

```
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 118; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;

  Query Match             88.9%;  Score 40;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
           |||||:|||
Db       1 ylsganlnl 9

RESULT   7
Y09525
ID   Y09525 standard; peptide; 9 AA.
XX
AC   Y09525;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist CAP-1.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
```

```
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 1; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             88.9%;  Score 40;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
           |||||:|||
Db       1 ylsganlnl 9

RESULT   8
Y46555
ID   Y46555 standard; Peptide; 10 AA.
XX
AC   Y46555;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #1166.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
```

```
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 76; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   10 AA;

  Query Match             88.9%;  Score 40;  DB 20;  Length 10;
  Best Local Similarity   88.9%;  Pred. No. 0.024;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
             |||||:|||
Db       2 ylsganlnl 10

RESULT   9
W86133
ID   W86133 standard; Protein; 107 AA.
XX
AC   W86133;
XX
DT   03-MAR-1999  (first entry)
XX
DE   Protein sequence of vaccine 2 708 Vl.
XX
KW   Non-immunogenic; epitope; T-cell; immunogenicity; immune system; SK;
KW   immunoglobulin; therapeutic; streptokinase; vaccine; 708.
XX
OS   Homo sapiens.
XX
PN   WO9852976-A1.
XX
PD   26-NOV-1998.
XX
PF   21-MAY-1998;   98WO-GB01473.
XX
PR   14-APR-1998;   98GB-0007751.
PR   21-MAY-1997;   97GB-0010480.
```

```
PR   31-JUL-1997;   97GB-0016197.
PR   28-NOV-1997;   97GB-0025270.
PR   02-DEC-1997;   97US-0067235.
XX
PA   (BIOV-) BIOVATION LTD.
XX
PI   Carr FJ;
XX
DR   WPI; 1999-045301/04.
XX
PT   Reducing immunogenicity of proteins - by modifying the amino acid
PT   sequence of the protein to eliminate potential epitopes for T-cells
PT   of a given species
XX
PS   Example 4; Fig 19; 77pp; English.
XX
CC   The invention relates to a method for the production of non-immunogenic
CC   proteins. The method comprises determining at least part of the amino
CC   acid sequence of the protein; (b) identifying in the amino acid sequence
CC   one or more potential epitopes for T-cells (T-cell epitopes) of the given
CC   species; and (c) modifying the amino acid sequence to eliminate at least
CC   one of the T-cell epitopes identified in step (b) thereby to eliminate or
CC   reduce the immunogenicity of the protein when exposed to the immune
CC   system of the given species. A method of analysing a pre-existing protein
CC   to predict the basis for immunogenic responses is also provided. The
CC   methods can be used particularly for reducing the immunogenicity of
CC   immunoglobulins or therapeutic proteins, e.g. Streptokinase (SK). The
CC   products can be used for diagnosis and therapy. The present sequence
CC   represents the protein sequence of vaccine 2 708 Vl.
XX
SQ   Sequence   107 AA;

  Query Match             88.9%;  Score 40;  DB 20;  Length 107;
  Best Local Similarity   88.9%;  Pred. No. 0.39;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
             |||||:|||
Db      89 ylsganlnl 97

RESULT  10
P93499
ID   P93499 standard; protein; 178 AA.
XX
AC   P93499;
XX
DT   08-MAY-1990  (first entry)
XX
DE   Sequence of carcinoembryonic antigen domain III.
XX
KW   Carcinoembryonic antigen; domain III; domain A; domain B.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..89
FT                   /note="domain A"
FT   Domain          90..178
FT                   /note="domain B"
XX
PN   EP343946-A.
XX
PD   29-NOV-1989.
XX
PF   24-MAY-1989;   89EP-0305232.
XX
PR   25-MAY-1988;   88US-0198289.
XX
PA   (CITY ) CITY OF HOPE.
XX
PI   Shively JE;
XX
DR   WPI; 1989-349991/48.
```

```
DR   N-PSDB; N92449.
XX
PT   Carcinoembryonic antigen fragments - used in assays to determine the
PT   presence and amt. of the antigen in samples also contg. related antigens.
XX
PS   Disclosure; page 4; 15pp; English.
XX
CC   CEA fragments can be used in assays to determine the presence and amt. of
CC   CEA in samples which also may contain related antigens including its
CC   normal cross-reacting antigen or the 128 kD antigen.
XX
SQ   Sequence   178 AA;

  Query Match             88.9%;  Score 40;  DB 10;  Length 178;
  Best Local Similarity   88.9%;  Pred. No. 0.71;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
             |||||:|||
Db     107 ylsganlnl 115

RESULT  11
R77436
ID   R77436 standard; Protein; 468 AA.
XX
AC   R77436;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-643) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..468
FT                   /note= "CEA (490-643)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 16; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   468 AA;

  Query Match             88.9%;  Score 40;  DB 16;  Length 468;
  Best Local Similarity   88.9%;  Pred. No. 2.2;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
             |||||:|||
Db     396 ylsganlnl 404

RESULT  12
R77435
ID   R77435 standard; Protein; 493 AA.
XX
AC   R77435;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-C-terminal) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..493
FT                   /note= "CEA (490-C-terminal)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 15; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
```

```
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   493 AA;

  Query Match             88.9%;  Score 40;  DB 16;  Length 493;
  Best Local Similarity   88.9%;  Pred. No. 2.3;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADLNL 9
              |||||:|||
Db      396 ylsganlnl 404

RESULT  13
R77437
ID   R77437 standard; Protein; 509 AA.
XX
AC   R77437;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-644)/BGP (391-430) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..469
FT                   /note= "CEA (490-644)"
FT   Protein         470..509
FT                   /note= "BGP (391-430)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 17; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
```

```
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   509 AA;

  Query Match             88.9%;  Score 40;  DB 16;  Length 509;
  Best Local Similarity   88.9%;  Pred. No. 2.4;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADLNL 9
              |||||:|||
Db      396 ylsganlnl 404

RESULT  14
R77438
ID   R77438 standard; Protein; 511 AA.
XX
AC   R77438;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-642)/BGP (387-430) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..467
FT                   /note= "CEA (490-642)"
FT   Protein         468..511
FT                   /note= "BGP (387-430)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 18; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
```

```
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   511 AA;


  Query Match             88.9%;  Score 40;  DB 16;  Length 511;
  Best Local Similarity   88.9%;  Pred. No. 2.4;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
             |||||:|||
Db     396 ylsganlnl 404


RESULT  15
R60619
ID   R60619 standard; Protein; 642 AA.
XX
AC   R60619;
XX
DT   10-MAY-1995  (first entry)
XX
DE   Carcinoembryonic antigen glycoprotein.
XX
KW   Carcinoembryonic antigen; CEA; neoplastic diseases.
XX
OS   Homo sapiens.
XX
PN   EP618292-A.
XX
PD   05-OCT-1994.
XX
PF   15-MAR-1994;   94EP-0103986.
XX
PR   25-MAR-1993;   93EP-0810214.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Mach J,  Pelegrin A,  Terskikh A;
XX
DR   WPI; 1994-304461/38.
DR   N-PSDB; Q71567.
XX
PT   Carcinoembryonic antigen (CEA) derivs - useful as reagents in
PT   immunoassay for diagnosis of neoplastic diseases
XX
PS   Claim 2; Page 15; 30pp; English.
XX
CC   Q71567 is the DNA sequence which encodes carcinoembryonic antigen (CEA)
CC   R60619. CEA is free from cross-reactive CEA-like antigens, it is
CC   antigenically indistinguishable from the solution form of CEA shed from
CC   tumour cells, and it is devoid of ethanolamine. R60619 can be used in a
CC   reagent composition for detecting neoplastic diseases in biological
CC   samples, or in an immunoassay process where it can specifically detect
CC   the presence of tumour cells in a biological sample e.g. blood.
XX
SQ   Sequence   642 AA;


  Query Match             88.9%;  Score 40;  DB 15;  Length 642;
  Best Local Similarity   88.9%;  Pred. No. 3.2;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
             |||||:|||
Db     571 ylsganlnl 579
```

```
Search completed: December 16, 2000, 00:51:14
Job time: 18774 sec
```

```
                          GenCore version 4.5
                 Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 04:22:12 ; Search time 111.26 Seconds
                                               (without alignments)
                                               7.553 Million cell updates/sec

Title:           US-09-529-121-2
Perfect score:   45
Sequence:        1 YLSGADLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters:      436

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       SPTREMBL_14:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                 %
Result          Query
  No.   Score   Match Length DB  ID                   Description
 -----------------------------------------------------------------------
    1     17    37.8      7  11  Q63480                Q63480 rattus norv
    2     16    35.6      8  13  P82079                P82079 limnodynast
    3     16    35.6      9  5  P82003                P82003 bombyx mori
    4     15    33.3      8  3  P87225                P87225 saccharomyc
    5     15    33.3      8  4  Q9UMC7                Q9umc7 homo sapien
    6     15    33.3      8  4  Q9UL56                Q9ul56 homo sapien
    7     14    31.1      8  2  Q9X3K1                Q9x3k1 prochloroco
    8     14    31.1      8  5  Q94695                Q94695 physarum po
    9     14    31.1      8  13  Q9PS69                Q9ps69 gallus gall
   10     14    31.1      9  1  Q50832                Q50832 methanococc
   11     14    31.1      9  4  Q16220                Q16220 homo sapien
   12     14    31.1      9  4  O95953                O95953 homo sapien
   13     14    31.1      9  6  Q9TRS0                Q9trs0 oryctolagus
   14     14    31.1      9  11  O35953                O35953 mus musculu
   15     13    28.9      7  11  O55184                O55184 rattus norv
   16     13    28.9      9  5  Q9TWD6                Q9twd6 leptinotars
   17     13    28.9      9  6  Q9TRW2                Q9trw2 oryctolagus
   18     12    26.7      8  2  P77556                P77556 escherichia
   19     12    26.7      8  4  Q15901                Q15901 homo sapien
```

```
   20     12    26.7      8  4  Q9UMH9                Q9umh9 homo sapien
   21     12    26.7      8  7  Q95213                Q95213 oryctolagus
   22     12    26.7      8  12  Q66807                Q66807 echovirus 2
   23     12    26.7      8  13  Q90498                Q90498 erythrura g
   24     12    26.7      8  13  Q91098                Q91098 manorina me
   25     12    26.7      9  2  Q57328                Q57328 aeromonas s
   26     12    26.7      9  2  Q44001                Q44001 aeromonas e
   27     12    26.7      9  2  Q44377                Q44377 aeromonas t
   28     12    26.7      9  2  Q44468                Q44468 aeromonas v
   29     12    26.7      9  2  Q43928                Q43928 aeromonas c
   30     12    26.7      9  2  Q9R635                Q9r635 chlamydia t
   31     12    26.7      9  4  Q9UCN5                Q9ucn5 homo sapien
   32     12    26.7      9  5  O96417                O96417 drosophila
   33     12    26.7      9  5  Q27396                Q27396 babesia bov
   34     12    26.7      9  6  Q28112                Q28112 bos taurus
   35     11    24.4      7  12  Q9YQ10                Q9yq10 porcine tra
   36     11    24.4      8  2  Q9RQ57                Q9rq57 buchnera ap
   37     11    24.4      8  2  Q9RQ49                Q9rq49 buchnera ap
   38     11    24.4      8  4  Q15889                Q15889 homo sapien
   39     11    24.4      8  4  Q9UCN4                Q9ucn4 homo sapien
   40     11    24.4      8  6  Q9XSY1                Q9xsy1 canis famil
   41     11    24.4      8  8  Q35792                Q35792 saccharomyc
   42     11    24.4      8  11  Q9QVF4                Q9qvf4 rattus sp.
   43     11    24.4      8  12  Q83332                Q83332 murine hepa
   44     11    24.4      8  13  Q90493                Q90493 eopsaltria
   45     11    24.4      9  2  Q9R7H9                Q9r7h9 haemophilus
   46     11    24.4      9  2  Q9R7E8                Q9r7e8 escherichia
   47     11    24.4      9  2  Q9R5M1                Q9r5m1 staphylococ
   48     11    24.4      9  4  P78484                P78484 homo sapien
   49     11    24.4      9  6  Q9XSL0                Q9xsl0 capra hircu
   50     11    24.4      9  13  Q92009                Q92009 gallus gall
   51     11    24.4      9  13  Q9PS68                Q9ps68 gallus gall
   52     10    22.2      7  2  O07354                O07354 synechococc
   53     10    22.2      7  4  Q15897                Q15897 homo sapien
   54     10    22.2      7  5  Q9VYN9                Q9vyn9 drosophila
   55     10    22.2      7  8  O98866                O98866 spinacia ol
   56     10    22.2      7  12  Q07624                Q07624 rous sarcom
   57     10    22.2      8  2  P72221                P72221 pseudomonas
   58     10    22.2      8  2  Q9ZIE9                Q9zie9 neisseria m
   59     10    22.2      8  2  Q9R7T2                Q9r7t2 escherichia
   60     10    22.2      8  2  Q9R4M3                Q9r4m3 enterococcu
   61     10    22.2      8  3  Q9URB9                Q9urb9 saccharomyc
   62     10    22.2      8  4  Q16428                Q16428 homo sapien
   63     10    22.2      8  4  Q9Y4J4                Q9y4j4 homo sapien
   64     10    22.2      8  4  Q9Y4J3                Q9y4j3 homo sapien
   65     10    22.2      8  4  Q9UDZ4                Q9udz4 homo sapien
   66     10    22.2      8  5  Q94623                Q94623 manduca sex
   67     10    22.2      8  5  Q9UB13                Q9ub13 albinaria h
   68     10    22.2      8  8  Q9XNP8                Q9xnp8 boophilus m
   69     10    22.2      8  10  Q40659                Q40659 oryza sativ
   70     10    22.2      8  11  O35835                O35835 rattus norv
   71     10    22.2      8  13  P82082                P82082 limnodynast
   72     10    22.2      8  13  P82083                P82083 limnodynast
   73     10    22.2      9  2  Q51349                Q51349 pseudomonas
   74     10    22.2      9  2  Q9R9C4                Q9r9c4 borrelia bu
   75     10    22.2      9  4  Q15999                Q15999 homo sapien
```

ALIGNMENTS

```
RESULT   1
Q63480
ID   Q63480      PRELIMINARY;      PRT;     7 AA.
AC   Q63480;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   TR4-NS ORPHAN RECEPTOR (FRAGMENT).
GN   TR4.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96198747.
RA   Yoshikawa T., Makino S., Gao X.M., Xing G.Q., Chuang D.M.,
RA   Detera-Wadleigh S.D.;
RT   "Splice variants of rat TR4 orphan receptor: differential expression
RT   of novel sequences in the 5'-untranslated region and C-terminal
RT   domain.";
RL   Endocrinology 137:1562-1571(1996).
DR   EMBL; U59125; AAB02827.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   7 AA;  758 MW;  672AA87864005350 CRC64;


  Query Match             37.8%;  Score 17;  DB 11;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        2 LSGADL 7
            : | ||
Db        2 IRGGDL 7


RESULT   2
P82079
ID   P82079      PRELIMINARY;      PRT;     8 AA.
AC   P82079;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   DYNASTIN 1.
OS   Limnodynastes interioris (Giant banjo frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Myobatrachidae;
OC   Limnodynastes.
RN   [1]
RP   SEQUENCE, AND MASS SPECTROMETRY.
RC   TISSUE=TIBIAL GLAND;
RA   Raftery M.J., Bradford A.M., Bowie J.H., Wallace J.C., Tyler M.J.;
RT   "Peptides from Australian frogs. The structure of the dynastins from
RT   the banjo frogs Limnodynastes interioras, Limnodynastes dumerilii and
RT   Limnodynastes terraereginae.";
RL   Aust. J. Chem. 46:833-842(1993).
CC   -!- MASS SPECTROMETRY: MW=729; METHOD=FAB.
KW   Amphibian skin.
SQ   SEQUENCE   8 AA;  729 MW;  7C28772865B72728 CRC64;


  Query Match             35.6%;  Score 16;  DB 13;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        2 LSGADL 7
            |||  |
Db        3 LSGLGL 8


RESULT   3
P82003
ID   P82003      PRELIMINARY;      PRT;     9 AA.
AC   P82003;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   PROTHORACICOSTATIC PEPTIDE (PTSP).
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
RN   [1]
RP   SEQUENCE.
RC   STRAIN=C145 X N140; TISSUE=BRAIN;
RA   Hua Y.-J., Tanaka Y., Nakamura K.;
RT   "Identification of a prothoracicostatic peptide (PTSP) from the larval
RT   brain of the silkworm, Bombyx mori.";
RL   J. Biol. Chem. 0:0-0(1999).
CC   -!- FUNCTION: INHIBITS ECDYSTEROIDOGENESIS BY PROTHORACIC GLAND IN THE
CC       SILKWORM.
CC   -!- DEVELOPMENTAL STAGE: EARLY FIFTH INSTAR.
KW   Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1090 MW;  3878C5B4472AB6C3 CRC64;


  Query Match             35.6%;  Score 16;  DB 5;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 DLN 8
            |||
Db        4 DLN 6


RESULT   4
P87225
ID   P87225      PRELIMINARY;      PRT;     8 AA.
AC   P87225;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GIN11 PROTEIN (FRAGMENT).
GN   GIN11.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wedler H., Wedler E., Scharfe M., Wambutt R.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z73169; CAA97518.2; -.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  1019 MW;  4E21A9C449D5B73B CRC64;


  Query Match             33.3%;  Score 15;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLS 3
            |||
Db        1 YLS 3


RESULT   5
Q9UMC7
ID   Q9UMC7      PRELIMINARY;      PRT;     8 AA.
AC   Q9UMC7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   SHMT PROTEIN (FRAGMENT).
GN   SHMT.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chave K.J., Snell K., Sanders P.G.;
RT   "Isolation and characterisation of human genomic sequences encoding
RT   cytosolic serine hydroxymethyltransferase.";
RL   Biochem. Soc. Trans. 25:53-53(1997).
DR   EMBL; Y14492; CAB54844.1; -.
```

```
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  868 MW;  7C205721E44AB5B8 CRC64;


  Query Match             33.3%;  Score 15;  DB 4;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         4 GADLNL 9
             |:| :|
Db         1 GSDNHL 6


RESULT   6
Q9UL56
ID   Q9UL56      PRELIMINARY;      PRT;     8 AA.
AC   Q9UL56;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MUTANT NADH-CYTOCHROME B5 REDUCTASE (EC 1.6.2.2) (FRAGMENT).
GN   DIA1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fukumaki Y., Higasa K.;
RT   "Two novel mutations in Thai patients with hereditary
RT   methemoglobinemia types I and II: a subtle amino acid change causes
RT   instability of NADH-cytochrome b5 reductase.";
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF061830; AAF06818.1; -.
KW   Oxidoreductase.
FT   NON_TER       1      1
FT   VARIANT       9      9       -> R.
SQ   SEQUENCE   8 AA;  888 MW;  76C5B73B5051F6D8 CRC64;


  Query Match             33.3%;  Score 15;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 YLS 3
             |||
Db         5 YLS 7


RESULT   7
Q9X3K1
ID   Q9X3K1      PRELIMINARY;      PRT;     8 AA.
AC   Q9X3K1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CYTOCHROME B (FRAGMENT).
GN   PETB.
OS   Prochlorococcus sp.
OC   Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC   Prochlorococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Urbach E., Chisholm S.W.;
RT   "Genetic diversity in Prochlorococcus populations flow cytometrically
RT   sorted from the Sargasso Sea and Gulf Stream.";
RL   Limnol. Oceanog. 43:1615-1630(1998).
DR   EMBL; AF070193; AAD23233.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  799 MW;  10376865B72866D3 CRC64;

  Query Match             31.1%;  Score 14;  DB 2;  Length 8;
```

```
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         2 LSG 4
             |||
Db         4 LSG 6


RESULT   8
Q94695
ID   Q94695      PRELIMINARY;      PRT;     8 AA.
AC   Q94695;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ACTIN (FRAGMENT).
GN   ARDC.
OS   Physarum polycephalum (Slime mold).
OC   Eukaryota; Myxogastria; Physarida; Physarum.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96182101.
RA   Benard M., Lagnel C., Pallotta D., Pierron G.;
RT   "Mapping of a replication origin within the promoter region of two
RT   unlinked, abundantly transcribed actin genes of Physarum
RT   polycephalum.";
RL   Mol. Cell. Biol. 16:968-976(1996).
DR   EMBL; M73459; AAB03706.1; -.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  878 MW;  F4C6C2CAAB187B16 CRC64;


  Query Match             31.1%;  Score 14;  DB 5;  Length 8;
  Best Local Similarity   33.3%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         2 LSGADL 7
               : | |:
Db         1 MEGEDV 6


RESULT   9
Q9PS69
ID   Q9PS69      PRELIMINARY;      PRT;     8 AA.
AC   Q9PS69;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   LOW DENSITY LIPOPROTEIN RECEPTOR-RELATED PROTEIN.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92011685.
RA   Stifani S., Barber D.L., Aebersold R., Steyrer E., Shen X., Nimpf J.,
RA   Schneider W.J.;
RT   "The laying hen expresses two different low density lipoprotein
RT   receptor-related proteins.";
RL   J. Biol. Chem. 266:19079-19087(1991).
SQ   SEQUENCE   8 AA;  846 MW;  C007272DD865BAAA CRC64;


  Query Match             31.1%;  Score 14;  DB 13;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         3 SGA 5
             |||
Db         3 SGA 5
```

```
RESULT  10
Q50832
ID   Q50832      PRELIMINARY;      PRT;     9 AA.
AC   Q50832;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1996 (TrEMBLrel. 01, Last annotation update)
DE   INTERGENIC AT-RICH DNA SEQUENCE (FRAGMENT).
OS   Methanococcus voltae.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85230552.
RA   Bollschweiler C., Kuehn R., Klein A.;
RT   "Non-repetitive AT-rich sequences are found in intergenic regions of
RT   Methanococcus voltae DNA.";
RL   EMBO J. 4:805-809(1985).
DR   EMBL; X02518; CAA26355.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1087 MW;  99ED005DC404405A CRC64;

  Query Match             31.1%;  Score 14;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 DLN 8
            |:|
Db        2 DIN 4

RESULT  11
Q16220
ID   Q16220      PRELIMINARY;      PRT;     9 AA.
AC   Q16220;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HGRP PROTEIN (FRAGMENT).
GN   HGRP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94320083.
RA   Nagalla S.R., Spindel E.R.;
RT   "Functional analysis of the 5'-flanking region of the human gastrin-
RT   releasing peptide gene in small cell lung carcinoma cell lines.";
RL   Cancer Res. 54:4461-4467(1994).
DR   EMBL; S73265; AAD14116.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1070 MW;  77FE37672B040864 CRC64;

  Query Match             31.1%;  Score 14;  DB 4;  Length 9;
  Best Local Similarity   37.5%;  Pred. No. 3e+05;
  Matches    3;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        2 LSGADLNL 9
             : | :| |
Db        1 MRGRELPL 8

RESULT  12
O95953
ID   O95953      PRELIMINARY;      PRT;     9 AA.
AC   O95953;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   GALACTOCEREBROSIDASE (EC 3.2.1.46) (FRAGMENT).
GN   GALC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Lulli L., Torchiana E., Finocchiaro G.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U77631; AAD15626.1; -.
KW   Hydrolase; Glycosidase.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1069 MW;  374E2AADC2C699C8 CRC64;

  Query Match             31.1%;  Score 14;  DB 4;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        5 ADL 7
            |||
Db        6 ADL 8

RESULT  13
Q9TRS0
ID   Q9TRS0      PRELIMINARY;      PRT;     9 AA.
AC   Q9TRS0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CALCYCLIN-ASSOCIATED PROTEIN, CAP50=CA2+/PHOSPHOLIPID-BINDING PROTEIN
DE   L-7 FRAGMENT.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92250478.
RA   Tokumitsu H., Mizutani A., Minami H., Kobayashi R., Hidaka H.;
RT   "A calcyclin-associated protein is a newly identified member of the
RT   Ca2+/phospholipid-binding proteins, annexin family.";
RL   J. Biol. Chem. 267:8919-8924(1992).
SQ   SEQUENCE   9 AA;  1010 MW;  64E419C44865B72B CRC64;

  Query Match             31.1%;  Score 14;  DB 6;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSG 4
            |||
Db        3 LSG 5

RESULT  14
O35953
ID   O35953      PRELIMINARY;      PRT;     9 AA.
AC   O35953;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   SODIUM CHANNEL, VOLTAGE-GATED, TYPE VIII, ALPHA POLYPEPTIDE
DE   (TRUNCATED VOLTAGE-GATED SODIUM CHANNEL ALPHA SUBUNIT) (FRAGMENT).
GN   SCN8A.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIII;
RX   MEDLINE; 97442476.
RA   Plummer N.W., McBurney M.W., Meisler M.H.;
```

```
RT   "Alternative splicing of the sodium channel SCN8A predicts a truncated
RT   two-domain protein in fetal brain and non-neuronal cells.";
RL   J. Biol. Chem. 272:24008-24015(1997).
DR   EMBL; U97672; AAB80914.1; -.
DR   MGD; MGI:103169; Scn8a.
KW   Ionic channel.
FT   NON_TER      1      1
SQ   SEQUENCE   9 AA;  898 MW;  22D92865B735B737 CRC64;

  Query Match             31.1%;  Score 14;  DB 11;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           |||
Db       5 LSG 7


RESULT  15
O55184
ID   O55184      PRELIMINARY;      PRT;     7 AA.
AC   O55184;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ORPHAN RECEPTOR TR4-NS (FRAGMENT).
GN   TR4.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96198747.
RA   Yoshikawa T., Makino S., Gao X.M., Xing G.Q., Chuang D.M.,
RA   Detera-Wadleigh S.D.;
RT   "Splice variants of rat TR4 orphan receptor: differential expression
RT   of novel sequences in the 5'-untranslated region and C-terminal
RT   domain.";
RL   Endocrinology 137:1562-1571(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96299786.
RA   Yoshikawa T., Dupont B.R., Leach R.J., Detera-Wadleigh S.D.;
RT   "New variants of the human and rat nuclear hormone receptor, TR4:
RT   expression and chromosomal localization of the human gene.";
RL   Genomics 35:361-366(1996).
DR   EMBL; U59454; AAB91433.1; -.
FT   NON_TER      1      1
SQ   SEQUENCE   7 AA;  663 MW;  6DDAA8787EB05350 CRC64;

  Query Match             28.9%;  Score 13;  DB 11;  Length 7;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LSGAD 6
           : | |
Db       2 ICGGD 6


RESULT  16
Q9TWD6
ID   Q9TWD6      PRELIMINARY;      PRT;     9 AA.
AC   Q9TWD6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   LED-NPF-1=NEUROPEPTIDE F-RELATED PEPTIDE.
OS   Leptinotarsa decemlineata (Colorado potato beetle).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC   Cucujiformia; Chrysomelidae; Chrysomelinae; Leptinotarsa.
```

```
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 96245438.
RA   Spittaels K., Verhaert P., Shaw C., Johnston R.N., Devreese B.,
RA   Van Beeumen J., De Loof A.;
RT   "Insect neuropeptide F (NPF)-related peptides: isolation from Colorado
RT   potato beetle (Leptinotarsa decemlineata) brain.";
RL   Insect Biochem. Mol. Biol. 26:375-382(1996).
SQ   SEQUENCE   9 AA;  1066 MW;  7E02340736C76864 CRC64;

  Query Match             28.9%;  Score 13;  DB 5;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 GADLNL 9
           |  | |
Db       3 GPQLRL 8


RESULT  17
Q9TRW2
ID   Q9TRW2      PRELIMINARY;      PRT;     9 AA.
AC   Q9TRW2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CALDESMON=PHOSPHORYLATION SITE.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 91378498.
RA   Ikebe M., Hornick T.;
RT   "Determination of the phosphorylation sites of smooth muscle caldesmon
RT   by protein kinase C.";
RL   Arch. Biochem. Biophys. 288:538-542(1991).
SQ   SEQUENCE   9 AA;  1018 MW;  8C901B10533735A5 CRC64;

  Query Match             28.9%;  Score 13;  DB 6;  Length 9;
  Best Local Similarity   33.3%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 GADLNL 9
           |: | :
Db       1 GSSLKI 6


RESULT  18
P77556
ID   P77556      PRELIMINARY;      PRT;     8 AA.
AC   P77556;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   TRAY (FRAGMENT).
GN   TRAY.
OS   Escherichia coli.
OG   Plasmid IncFII R1.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ECOR11;
RX   MEDLINE; 96400908.
RA   Boyd E.F., Hill C.W., Rich S.M., Hartl D.L.;
RT   "Mosaic structure of plasmids from natural populations of Escherichia
RT   coli.";
RL   Genetics 143:1091-1100(1996).
DR   EMBL; U50661; AAC44245.1; -.
DR   EMBL; U50650; AAC44234.1; -.
```

```
DR   EMBL; U50651; AAC44235.1; -.
DR   EMBL; U50652; AAC44236.1; -.
DR   EMBL; U50653; AAC44237.1; -.
DR   EMBL; U50654; AAC44238.1; -.
DR   EMBL; U50655; AAC44239.1; -.
DR   EMBL; U50656; AAC44240.1; -.
DR   EMBL; U50657; AAC44241.1; -.
DR   EMBL; U50658; AAC44242.1; -.
DR   EMBL; U50659; AAC44243.1; -.
DR   EMBL; U50660; AAC44244.1; -.
KW   Plasmid.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  834 MW;  D335A5B0544735A1 CRC64;

  Query Match             26.7%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LNL 9
           ||:
Db       3 LNI 5

RESULT  19
Q15901
ID   Q15901      PRELIMINARY;      PRT;     8 AA.
AC   Q15901;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   (CLONE XP7B11B) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   Lee C.C., Yazdani A., Wehnert M., Bailey J., Couch L., Xiong M.,
RA   Coolbaugh M.I., Chinault C.A., Baldini A., Lindsay E.A., Zhao Z.Y.,
RA   Caskey C.T.H.;
RL   Hum. Mol. Genet. 0:0-0(0).
DR   EMBL; L32080; AAA73891.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  860 MW;  37D72878676729CB CRC64;

  Query Match             26.7%;  Score 12;  DB 4;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           :| |
Db       2 FLPG 5

RESULT  20
Q9UMH9
ID   Q9UMH9      PRELIMINARY;      PRT;     8 AA.
AC   Q9UMH9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RHCE PROTEIN (FRAGMENT).
GN   RHCE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RA   Matassi G., Cherif-Zahar B., Mouro I., Cartron J.P.;
```

```
RT   "Characterization of the recombination hot spot involved in the
RT   genomic rearrangement leading to the hybrid D-CE-D gene in the DVI
RT   phenotype.";
RL   Am. J. Hum. Genet. 60:808-817(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE; 90349591.
RA   Cherif-Zahar B., Bloy C., Le Van Kim C., Blanchard D., Bailly P.,
RA   Hermand P., Salmon C., Cartron J.-P., Colin Y.;
RT   "Molecular cloning and protein structure of a human blood group Rh
RT   polypeptide.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:6243-6247(1990).
DR   EMBL; Z97030; CAB09726.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  1049 MW;  C007244691FB5AB1 CRC64;

  Query Match             26.7%;  Score 12;  DB 4;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LNL 9
           :||
Db       5 MNL 7

RESULT  21
Q95213
ID   Q95213      PRELIMINARY;      PRT;     8 AA.
AC   Q95213;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   GERMLINE DH (DF) GENE (FRAGMENT).
GN   DF.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F-I/RGM;
RA   Mage R.G., Chen H.T., Alexander C.B., Chen F.F.;
RL   Mol. Immunol. 0:0-0(0).
DR   EMBL; U62585; AAB18735.1; -.
KW   MHC.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  845 MW;  5CA861B5AB58677B CRC64;

  Query Match             26.7%;  Score 12;  DB 7;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
             | :|
Db       4 YSTG 7

RESULT  22
Q66807
ID   Q66807      PRELIMINARY;      PRT;     8 AA.
AC   Q66807;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   5'UTR IN IRES SEQUENCES (ISOLATE TH222) (FRAGMENT).
OS   Echovirus 25.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Enterovirus.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=TH222;
RA   Bailly J.L., Borman A.M., Peigue-Lafeuille H., Kean K.M.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X90724; CAA62259.1; -.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  821 MW;  EFC1B5A2D6DDD876 CRC64;

  Query Match             26.7%;  Score 12;  DB 12;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GADLN 8
           || ::
Db       2 GAQVS 6

RESULT  23
Q90498
ID   Q90498      PRELIMINARY;      PRT;     8 AA.
AC   Q90498;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   MYOGLOBIN (FRAGMENT).
OS   Erythrura gouldiae (Gouldian finch).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Passeridae; Erythrura.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EGG1;
RX   MEDLINE; 98208049.
RA   Heslewood M.M., Elphinstone M.S., Tidemann S.C., Baverstock P.R.;
RT   "Myoglobin intron variation in the Gouldian Finch Erythrura gouldiae
RT   assessed by temperature gradient gel electrophoresis.";
RL   Electrophoresis 19:142-151(1998).
DR   EMBL; U40496; AAC60363.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  890 MW;  C6C1F2C865B046DE CRC64;

  Query Match             26.7%;  Score 12;  DB 13;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       3 ISG 5

RESULT  24
Q91098
ID   Q91098      PRELIMINARY;      PRT;     8 AA.
AC   Q91098;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   MYOGLOBIN (FRAGMENT).
OS   Manorina melanocephala (noisy miner).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Meliphagidae; Manorina.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D02;
RX   MEDLINE; 98208049.
RA   Heslewood M.M., Elphinstone M.S., Tidemann S.C., Baverstock P.R.;
RT   "Myoglobin intron variation in the Gouldian Finch Erythrura gouldiae
RT   assessed by temperature gradient gel electrophoresis.";
RL   Electrophoresis 19:142-151(1998).
DR   EMBL; U40497; AAC60364.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  890 MW;  C6C1F2C865B046DE CRC64;

  Query Match             26.7%;  Score 12;  DB 13;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       3 ISG 5

RESULT  25
Q57328
ID   Q57328      PRELIMINARY;      PRT;     9 AA.
AC   Q57328;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (STRAIN 132) (FRAGMENT).
GN   EXEF'.
OS   Aeromonas sobria.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=132, FROM DR M. ALTWEGG UNIV. ZURICH CULTRE COLLECTION;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89466; CAA61647.1; -.
DR   EMBL; X89465; CAA61645.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1077 MW;  79E852C69444472B CRC64;

  Query Match             26.7%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DLN 8
           :||
Db       2 ELN 4

RESULT  26
Q44001
ID   Q44001      PRELIMINARY;      PRT;     9 AA.
AC   Q44001;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (STRAIN CDC 9179-79) (FRAGMENT).
GN   EXEF'.
OS   Aeromonas eucrenophila.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 9179-79;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89461; CAA61637.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1059 MW;  785852C69444472B CRC64;
```

```
  Query Match             26.7%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 DLN 8
            :||
Db        2 ELN 4


RESULT  27
Q44377
ID   Q44377      PRELIMINARY;      PRT;     9 AA.
AC   Q44377;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (STRAIN ATCC 49659) (FRAGMENT).
GN   EXEF'.
OS   Aeromonas trota.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 49659;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89468; CAA61651.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1059 MW;  785852C69444472B CRC64;


  Query Match             26.7%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 DLN 8
            :||
Db        2 ELN 4


RESULT  28
Q44468
ID   Q44468      PRELIMINARY;      PRT;     9 AA.
AC   Q44468;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (STRAIN CDC 1306-83) (FRAGMENT).
GN   EXEF'.
OS   Aeromonas veronii.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1306-83;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89457; CAA61629.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1059 MW;  785852C69444472B CRC64;


  Query Match             26.7%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 DLN 8
```

```
            :||
Db        2 ELN 4


RESULT  29
Q43928
ID   Q43928      PRELIMINARY;      PRT;     9 AA.
AC   Q43928; Q43918; Q43920; Q43921;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DNA FOR EXEF-EXEG INTERGENIC REGION (FRAGMENT).
GN   EXEF'.
OS   Aeromonas caviae.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC   Aeromonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VARIOUS STRAINS;
RX   MEDLINE; 97089747.
RA   Karlyshev A.V., Macintyre S.;
RT   "Study of the intergenic exeF-exeG region and its application as a
RT   simple preliminary test for Aeromonas spp.";
RL   FEMS Microbiol. Lett. 137:37-44(1996).
DR   EMBL; X89464; CAA61643.1; -.
DR   EMBL; X89462; CAA61639.1; -.
DR   EMBL; X89460; CAA61635.1; -.
DR   EMBL; X89463; CAA61641.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  1059 MW;  785852C69444472B CRC64;


  Query Match             26.7%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 DLN 8
            :||
Db        2 ELN 4


RESULT  30
Q9R635
ID   Q9R635      PRELIMINARY;      PRT;     9 AA.
AC   Q9R635;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR OUTER MEMBRANE PROTEIN VARIABLE DOMAIN IV, MOMP VD IV.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92040090.
RA   Peterson E.M., Cheng X., Markoff B.A., Fielder T.J., de la Maza L.M.;
RT   "Functional and structural mapping of Chlamydia trachomatis species-
RT   specific major outer membrane protein epitopes by use of neutralizing
RT   monoclonal antibodies.";
RL   Infect. Immun. 59:4147-4153(1991).
SQ   SEQUENCE   9 AA;  976 MW;  9C61B041B7645361 CRC64;


  Query Match             26.7%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSG 4
            :||
Db        7 ISG 9


RESULT  31
Q9UCN5
```

```
ID   Q9UCN5      PRELIMINARY;      PRT;     9 AA.
AC   Q9UCN5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CELL-SURFACE HEPARIN/HEPARANSULFATE-BINDING PROTEIN PEPTIDE 2
DE   (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92291065.
RA   Raboudi N., Julian J., Rohde L.H., Carson D.D.;
RT   "Identification of cell-surface heparin/heparan sulfate-binding
RT   proteins of a human uterine epithelial cell line (RL95).";
RL   J. Biol. Chem. 267:11930-11939(1992).
SQ   SEQUENCE   9 AA;  1008 MW;  CB56D0544732C732 CRC64;


  Query Match             26.7%;  Score 12;  DB 4;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LNL 9
           ||:
Db       5 LNI 7


RESULT  32
O96417
ID   O96417      PRELIMINARY;      PRT;     9 AA.
AC   O96417;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   SXL E1 FORM (FRAGMENT).
GN   SXL.
OS   Drosophila virilis (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98337843.
RA   Erickson J.W., Cline T.W.;
RT   "Key aspects of the primary sex determination mechanism are conserved
RT   across the genus Drosophila.";
RL   Development 125:3259-3268(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Bell M., Cline T.W.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF046045; AAC97605.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1089 MW;  9A8BD1AAA9C449CA CRC64;


  Query Match             26.7%;  Score 12;  DB 5;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       6 DLN 8
           | |
Db       2 DFN 4


RESULT  33
Q27396
ID   Q27396      PRELIMINARY;      PRT;     9 AA.
AC   Q27396;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   RHOPTRY ASSOCIATED PROTEIN 1.
GN   RAP-1.
OS   Babesia bovis.
OC   Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MO7;
RA   Suarez C.E., Palmer G.H., Hines S.A., McElwain T.F.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L77326; AAA96415.1; -.
SQ   SEQUENCE   9 AA;  931 MW;  6D5BD2C865B05044 CRC64;


  Query Match             26.7%;  Score 12;  DB 5;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       4 ISG 6


RESULT  34
Q28112
ID   Q28112      PRELIMINARY;      PRT;     9 AA.
AC   Q28112;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   6-PHOSPHOFRUCTO-2-KINASE/FRUCTOSE-2, 6-BISPHOSPHATASE (FRAGMENT).
GN   GENE B.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93387464.
RA   Vidal H., Crepin K.M., Rider M.H., Hue L., Rousseau G.G.;
RT   "Cloning and expression of novel isoforms of 6-phosphofructo-2-
RT   kinase/fructose-2,6-bisphosphatase from bovine heart.";
RL   FEBS Lett. 330:329-333(1993).
DR   EMBL; X74564; CAA52652.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  837 MW;  859CA5BDC7644865 CRC64;


  Query Match             26.7%;  Score 12;  DB 6;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       1 MSG 3


RESULT  35
Q9YQ10
ID   Q9YQ10      PRELIMINARY;      PRT;     7 AA.
AC   Q9YQ10;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HYPOTHETICAL FUSION PROTEIN.
OS   porcine transmissible gastroenteritis virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99099045.
RA   Izeta A., Smerdou C., Alonso S., Penzes Z., Mendez A., Plana-Duran J.,
RA   Enjuanes L.;
```

```
RT   "Replication and packaging of transmissible gastroenteritis
RT   coronavirus-derived synthetic minigenomes.";
RL   J. Virol. 73:1535-1545(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95159435.
RA   Eleouet J., Rasschaert D., Lambert P., Levy L., Vende P., Laude H.;
RT   "Complete sequence (20 kilobases) of the polyprotein-encoding gene 1
RT   of transmissible gastroenteritis virus.";
RL   Virology 206:817-822(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88078100.
RA   Rasschaert D., Gelfi J., Laude H.;
RT   "Enteric coronavirus TGEV: partial sequence of the genomic RNA, its
RT   organization and expression.";
RL   Biochimie 69:591-600(1987).
DR   EMBL; AJ011482; CAA09625.1; -.
SQ   SEQUENCE   7 AA;  927 MW;  69D6D7273B5726F0 CRC64;


  Query Match             24.4%;  Score 11;  DB 12;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YL 2
           ||
Db       3 YL 4


Search completed: December 16, 2000, 04:22:12
Job time: 4606 sec
```

```
                    GenCore version 4.5
              Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 04:23:31 ; Search time 68.03 Seconds
                                              (without alignments)
                                              4.227 Million cell updates/sec

Title:          US-09-529-121-2
Perfect score:  45
Sequence:       1 YLSGADLNL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters:      213

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :      SwissProt_39:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                     %
Result               Query
  No.     Score   Match Length DB  ID                      Description
-------------------------------------------------------------------------------
    1       17    37.8      8  1  CPD1_ENTFA              P13269 enterococcu
    2       16    35.6      8  1  LCK8_LEUMA              P19990 leucophaea
    3       16    35.6      9  1  LMIP_LOCMI              P31799 locusta mig
    4       15    33.3      9  1  FAR6_CALVO              P41861 calliphora
    5       14    31.1      8  1  ALL5_CYDPO              P82156 cydia pomon
    6       14    31.1      9  1  ISOT_CYPCA              P42993 cyprinus ca
    7       13    28.9      8  1  PLP_BRANA               P81707 brassica na
    8       13    28.9      9  1  DSIP_RABIT              P01158 oryctolagus
    9       13    28.9      9  1  FAR2_PANRE              P41873 panagrellus
   10       13    28.9      9  1  OXYA_SQUAC              P42999 squalus aca
   11       13    28.9      9  1  OXYT_RAJCL              P42994 raja clavat
   12       12    26.7      8  1  CCKN_MACEU              P30369 macropus eu
   13       12    26.7      9  1  FARD_CALVO              P41868 calliphora
   14       12    26.7      9  1  OXYT_RABIT              P32878 oryctolagus
   15       12    26.7      9  1  PGLR_DIAAB              P81179 diaprepes a
   16       11    24.4      4  1  FAR3_HIRME              P42562 hirudo medi
   17       11    24.4      5  1  PRCT_PERAM              P01373 periplaneta
   18       11    24.4      6  1  CIP2_MYTED              P13737 mytilus edu
   19       11    24.4      7  1  FAR2_ASCSU              P31890 ascaris suu
   20       11    24.4      7  1  GFRP_MOUSE              P99025 mus musculu
   21       11    24.4      7  1  LANC_CARUI              P36960 carnobacter
   22       11    24.4      7  1  UF03_MOUSE              P38641 mus musculu
   23       11    24.4      8  1  CAD1_ENTFA              P13268 enterococcu
   24       11    24.4      8  1  FAR3_HOMAM              P41486 homarus ame
   25       11    24.4      8  1  FAR8_CALVO              P41863 calliphora
   26       11    24.4      9  1  D1_NEPNO                P24816 nephrops no
   27       11    24.4      9  1  FAR5_CALVO              P41860 calliphora
   28       11    24.4      9  1  FAR7_CALVO              P41862 calliphora
   29       11    24.4      9  1  MOSF_CLYJA              P19853 clypeaster
   30       10    22.2      4  1  ACH1_ACHFU              P35904 achatina fu
   31       10    22.2      5  1  UXA4_CHLTR              P38005 chlamydia t
   32       10    22.2      6  1  TMOF_SARBU              P41495 sarcophaga
   33       10    22.2      6  1  TRPI_PSEPU              P36414 pseudomonas
```

```
   34       10    22.2      7  1  ALL2_CARMA              P81805 carcinus ma
   35       10    22.2      7  1  ALL3_CARMA              P81806 carcinus ma
   36       10    22.2      7  1  ALL4_CARMA              P81807 carcinus ma
   37       10    22.2      7  1  ALL5_CARMA              P81808 carcinus ma
   38       10    22.2      7  1  ALL7_CYDPO              P82158 cydia pomon
   39       10    22.2      7  1  FAR2_PROCL              P38498 procambarus
   40       10    22.2      7  1  UN06_PINPS              P81675 pinus pinas
   41       10    22.2      8  1  ACT_CARMA               P80709 carcinus ma
   42       10    22.2      8  1  AKH_MELML               P25423 melolontha
   43       10    22.2      8  1  AL12_CARMA              P81815 carcinus ma
   44       10    22.2      8  1  AL17_CARMA              P81820 carcinus ma
   45       10    22.2      8  1  ALL7_CARMA              P81809 carcinus ma
   46       10    22.2      8  1  ALL8_CARMA              P81811 carcinus ma
   47       10    22.2      8  1  ALL9_CARMA              P81812 carcinus ma
   48       10    22.2      8  1  FAR1_PANRE              P41872 panagrellus
   49       10    22.2      8  1  GLUR_HUMAN              P02729 homo sapien
   50       10    22.2      8  1  HTF_TENMO               P25419 tenebrio mo
   51       10    22.2      8  1  LCK5_LEUMA              P19987 leucophaea
   52       10    22.2      8  1  LPMS_STAEP              P23211 staphylococ
   53       10    22.2      8  1  RPCH_PANBO              P08939 pandalus bo
   54       10    22.2      8  1  UF06_MOUSE              P38644 mus musculu
   55       10    22.2      8  1  VGLG_HSV2B              P81780 herpes simp
   56       10    22.2      9  1  AL10_CARMA              P81813 carcinus ma
   57       10    22.2      9  1  AL11_CARMA              P81814 carcinus ma
   58       10    22.2      9  1  FAR5_ASCSU              P43170 ascaris suu
   59       10    22.2      9  1  FIBB_MACFU              P19345 macaca fusc
   60       10    22.2      9  1  MOSH_CLYJA              P19852 clypeaster
   61       10    22.2      9  1  OXYA_SCYCA              P42996 scyliorhinu
   62       10    22.2      9  1  OXYT_BUFRE              P42995 bufo regula
   63       10    22.2      9  1  TKL1_LOCMI              P16223 locusta mig
   64       10    22.2      9  1  TRP4_LEUMA              P81736 leucophaea
   65       10    22.2      9  1  ULAH_HUMAN              P31934 homo sapien
   66       10    22.2      9  1  UPA6_HUMAN              P30092 homo sapien
   67        9    20.0      4  1  EOSI_HUMAN              P02731 homo sapien
   68        9    20.0      4  1  FAR4_HIRME              P42563 hirudo medi
   69        9    20.0      5  1  AL14_CARMA              P81817 carcinus ma
   70        9    20.0      5  1  FARP_ARTTR              P41853 artioposthi
   71        9    20.0      7  1  FAR3_HAECO              P81298 haemonchus
   72        9    20.0      7  1  FAR3_PANRE              P41874 panagrellus
   73        9    20.0      7  1  FAR5_HIRME              P42564 hirudo medi
   74        9    20.0      8  1  AL15_CARMA              P81818 carcinus ma
   75        9    20.0      8  1  AL16_CARMA              P81819 carcinus ma


                               ALIGNMENTS


RESULT   1
CPD1_ENTFA
ID   CPD1_ENTFA     STANDARD;      PRT;     8 AA.
AC   P13269;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   SEX PHEROMONE CPD1.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 85040388.
RA   Suzuki A., Mori M., Sagakami Y., Isogai A., Fujino M., Kitada C.,
RA   Craig R.A., Clewell D.B.;
RT   "Isolation and structure of bacterial sex pheromone, cPD1.";
RL   Science 226:849-850(1984).
CC   -!- FUNCTION: CPD1 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE
CC       BACTERIOCIN PLASMID PPD1.
KW   Pheromone.
SQ   SEQUENCE   8 AA;  913 MW;  8665B729C682C729 CRC64;


  Query Match            37.8%;  Score 17;  DB 1;  Length 8;
  Best Local Similarity   75.0%;  Pred. No. 8.8e+04;
```

```
  Matches      3;  Conservative     1;  Mismatches     0;  Indels     0;  Gaps     0;

Qy       1 YLSG 4
           :|||
Db       5 FLSG 8

RESULT   2
LCK8_LEUMA
ID   LCK8_LEUMA     STANDARD;      PRT;     8 AA.
AC   P19990;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   LEUCOKININ VIII (L-VIII).
OS   Leucophaea maderae (Madeira cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blaberoidea; Blaberidae; Leucophaea.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=HEAD;
RA   Holman G.M., Cook B.J., Nachman R.J.;
RT   "Isolation, primary structure and synthesis of leucokinins VII and
RT   VIII: the final members of this new family of cephalomyotropic
RT   peptides isolated from head extracts of Leucophaea maderae.";
RL   Comp. Biochem. Physiol. 88C:31-34(1987).
CC   -!- FUNCTION: THIS CEPHALOMYOTROPIC PEPTIDE STIMULATES CONTRACTILE
CC       ACTIVITY OF COCKROACH PROTODEUM (HINDGUT).
CC   -!- SIMILARITY: TO THE OTHER LEUCOKININS.
DR   PIR; JS0318; JS0318.
KW   Neuropeptide; Amidation.
FT   MOD_RES      8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  902 MW;  736365AB59CAADD8 CRC64;

  Query Match             35.6%;  Score 16;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches      3;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy       4 GAD 6
           |||
Db       1 GAD 3

RESULT   3
LMIP_LOCMI
ID   LMIP_LOCMI     STANDARD;      PRT;     9 AA.
AC   P31799;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-OCT-1993 (Rel. 27, Last annotation update)
DE   LOCUSTAMYOINHIBITING PEPTIDE (LOM-MIP).
OS   Locusta migratoria (Migratory locust).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC   Acridomorpha; Acridoidea; Acrididae; Locusta.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92179466.
RA   Schoofs L., Holman G.M., Hayes T.K., Nachman R.J., de Loof A.;
RT   "Isolation, identification and synthesis of locustamyoinhibiting
RT   peptide (LOM-MIP), a novel biologically active neuropeptide from
RT   Locusta migratoria.";
RL   Regul. Pept. 36:111-119(1991).
CC   -!- FUNCTION: SUPPRESSES SPONTANEOUS CONTRACTIONS OF THE HINDGUT AND
CC       OVIDUCT.
CC   -!- TISSUE SPECIFICITY: NEURONS LOCATED IN TWO VENTRAL CELL CLUSTERS
CC       IN THE SUBOESOPHAGEAL GANGLION.
DR   PIR; A60065; AKLQIM.
KW   Amidation; Neuropeptide.
FT   MOD_RES      9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1060 MW;  387D7DD4472AB6C3 CRC64;
```

```
  Query Match             35.6%;  Score 16;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches      3;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy       6 DLN 8
           |||
Db       4 DLN 6

RESULT   4
FAR6_CALVO
ID   FAR6_CALVO     STANDARD;      PRT;     9 AA.
AC   P41861;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 6.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; F41978; F41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES      9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1058 MW;  96D10699CAB6D865 CRC64;

  Query Match             33.3%;  Score 15;  DB 1;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 8.8e+04;
  Matches      3;  Conservative     0;  Mismatches     1;  Indels     0;  Gaps     0;

Qy       3 SGAD 6
           || |
Db       2 SGQD 5

RESULT   5
ALL5_CYDPO
ID   ALL5_CYDPO     STANDARD;      PRT;     8 AA.
AC   P82156;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYDIASTATIN 5.
OS   Cydia pomonella (Codling moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Tortricoidea; Tortricidae; Olethreutinae; Cydia.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LARVA;
RX   MEDLINE; 98054539.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
RA   Davey M., East P.D., Thorpe A.;
RT   "Lepidopteran peptides of the allatostatin superfamily.";
RL   Peptides 18:1301-1309(1997).
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES      8      8       AMIDATION.
```

```
SQ   SEQUENCE   8 AA;  898 MW;  922879CABB58640D CRC64;

  Query Match             31.1%;  Score 14;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    3;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        4 GADLNL 9
            | |  |
Db        3 GYDFGL 8


RESULT   6
ISOT_CYPCA
ID   ISOT_CYPCA     STANDARD;      PRT;     9 AA.
AC   P42993;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ISOTOCIN.
OS   Cyprinus carpio (Common carp).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PITUITARY;
RA   Acher R., Chauvet J., Chauvet M.-T., Crepy D.;
RT   "Characterization of neurohypophyseal hormones from a fresh water bony
RT   fish, the carp (Cyprinus carpio). Comparison with hormones from sea
RT   water bony fishs.";
RL   Comp. Biochem. Physiol. 14:245-254(1965).
CC   -!- FUNCTION: ANTIDIURETIC HORMONE.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   PIR; A61364; A61364.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  969 MW;  17FF476EB455B04B CRC64;


  Query Match             31.1%;  Score 14;  DB 1;  Length 9;
  Best Local Similarity   28.6%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        1 YLSGADL 7
            |:|   :
Db        2 YISNCPI 8


RESULT   7
PLP_BRANA
ID   PLP_BRANA      STANDARD;      PRT;     8 AA.
AC   P81707;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PLASTIDIAL LIPID-ASSOCIATED PROTEIN (FRAGMENT).
OS   Brassica napus (Rape).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Brassica.
RN   [1]
RP   SEQUENCE.
RC   STRAIN=CV. TOPAZ; TISSUE=TAPETUM;
RX   MEDLINE; 99349136.
RA   Hernandez-Pinzon I., Ross J.H.E., Barnes K.A., Damant A.P.,
RA   Murphy D.J.;
RT   "Composition and role of tapetal lipid bodies in the biogenesis of the
RT   pollen coat of Brassica napus.";
RL   Planta 208:588-598(1999).
CC   -!- FUNCTION: MAY PLAY A STRUCTURAL ROLE IN THE ELAIOPLAST, A TAPETUM-
CC       SPECIFIC PLASTIDIAL LIPID ORGANELLE.
CC   -!- TISSUE SPECIFICITY: TAPETUM OF ANTHERS.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  989 MW;  9D7B1AA452CAA042 CRC64;


  Query Match             28.9%;  Score 13;  DB 1;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 DLN 8
            |:|
Db        3 DVN 5


RESULT   8
DSIP_RABIT
ID   DSIP_RABIT     STANDARD;      PRT;     9 AA.
AC   P01158;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   DELTA SLEEP-INDUCING PEPTIDE (DSIP).
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 77185324.
RA   Monnier M., Dudler L., Gachter R., Maier P.F., Tobler H.J.,
RA   Schoenenberger G.A.;
RT   "The delta sleep inducing peptide (DSIP). Comparative properties of
RT   the original and synthetic nonapeptide.";
RL   Experientia 33:548-552(1977).
RN   [2]
RP   SEQUENCE, AND SYNTHESIS.
RX   MEDLINE; 79054421.
RA   Schoenenberger G.A., Maier P.F., Tobler H.J., Wilson K., Monnier M.;
RT   "The delta EEG (sleep)-inducing peptide (DSIP). XI. Amino-acid
RT   analysis, sequence, synthesis and activity of the nonapeptide.";
RL   Pflugers Arch. 376:119-129(1978).
CC   -!- FUNCTION: WHEN INFUSED INTO THE MESODIENCEPHALIC VENTRICLE OF
CC       RECIPIENT RABBITS INDUCES SPINDLE AND DELTA EEG ACTIVITY AND
CC       REDUCED MOTOR ACTIVITIES.
CC   -!- MISCELLANEOUS: THIS PEPTIDE WAS OBTAINED FROM DIALYSATES OF
CC       OCCIPITAL VENOUS SINUS BLOOD FROM RABBITS KEPT ASLEEP BY ELECTRIC
CC       STIMULATION OF THE THALAMUS.
DR   PIR; A01422; QDRB.
SQ   SEQUENCE   9 AA;  849 MW;  DDD365BDDAA8787D CRC64;


  Query Match             28.9%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        3 SGAD 6
            :| |
Db        2 AGGD 5


RESULT   9
FAR2_PANRE
ID   FAR2_PANRE     STANDARD;      PRT;     9 AA.
AC   P41873;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE PF2 (SADPNFLRF-AMIDE).
OS   Panagrellus redivivus.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
OC   Panagrolaimoidea; Panagrolaimidae; Panagrellus.
```

```
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 93027659.
RA   Geary T.G., Price D.A., Bowman J.W., Winterrowd C.A., Mackenzie C.D.,
RA   Garrison R.D., Williams J.F., Friedman A.R.;
RT   "Two FMRFamide-like peptides from the free-living nematode
RT   Panagrellus redivivus.";
RL   Peptides 13:209-214(1992).
CC   -!- FUNCTION: MYOACTIVE.
CC   -!- TISSUE SPECIFICITY: NERVE CORDS AND PAIRED GROUPS OF CELLS LOCATED
CC       CAUDALLY TO THE BASE OF THE PHARYNX.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1066 MW;  DA0B0729C4576AAD CRC64;


  Query Match             28.9%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 8.8e+04;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        5 ADLN 8
            || |
Db        2 ADPN 5


RESULT  10
OXYA_SQUAC
ID   OXYA_SQUAC     STANDARD;      PRT;     9 AA.
AC   P42999;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ASPARTOCIN (ASPARGTOCIN).
OS   Squalus acanthias (Spiny dogfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Squaloidei; Squalidae; Squalus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 73031727.
RA   Acher R., Chauvet J., Chauvet M.-T.;
RT   "Phylogeny of the neurohypophysial hormones. Two new active peptides
RT   isolated from a cartilaginous fish, Squalus acanthias.";
RL   Eur. J. Biochem. 29:12-19(1972).
RN   [2]
RP   SEQUENCE.
RX   MEDLINE; 72128038.
RA   Acher R., Chauvet J., Chauvet M.-T., Fontaine M.;
RT   "Identification of 2 new neurohypophyseal hormones, valitocin (Val8-
RT   oxytocin) and aspartocin (Asn4-oxytocin) in a selachian fish, the
RT   spiny dog-fish (Squalus acanthias).";
RL   C. R. Acad. Sci., D, Sci. Nat. 274:313-316(1972).
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  996 MW;  17F8376EB444404B CRC64;


  Query Match             28.9%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   28.6%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        1 YLSGADL 7
            |::   |
Db        2 YINNCPL 8


RESULT  11
OXYT_RAJCL
ID   OXYT_RAJCL     STANDARD;      PRT;     9 AA.
AC   P42994;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   GLUMITOCIN.
OS   Raja clavata (Thornback ray).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Hypnosqualea; Pristiorajea; Batoidea;
OC   Rajiformes; Rajidae; Raja.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 66123415.
RA   Acher R., Chauvet J., Chauvet M.-T., Crepy D.;
RT   "Phylogeny of neurophyophyseal peptides: isolation of a new hormone,
RT   glumitocin (Ser 4-Gln 8-ocytocin) present in a cartilaginous fish,
RT   the ray (Raia clavata).";
RL   Biochim. Biophys. Acta 107:393-396(1965).
CC   -!- FUNCTION: ANTIDIURETIC HORMONE.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  984 MW;  17E9C76EB455B04B CRC64;


  Query Match             28.9%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLS 3
            |:|
Db        2 YIS 4


RESULT  12
CCKN_MACEU
ID   CCKN_MACEU     STANDARD;      PRT;     8 AA.
AC   P30369;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   CHOLECYSTOKININ (CCK).
GN   CCK.
OS   Macropus eugenii (Tammar wallaby), and
OS   Dasyurus viverrinus (Southeastern quoll).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=BRAIN;
RX   MEDLINE; 88234141.
RA   Fan Z.W., Eng J., Shaw G., Yalow R.S.;
RT   "Cholecystokinin octapeptide purified from brains of Australian
RT   marsupials.";
RL   Peptides 9:429-431(1988).
CC   -!- FUNCTION: THIS PEPTIDE HORMONE INDUCES GALL BLADDER CONTRACTION
CC       AND THE RELEASE OF PANCREATIC ENZYMES IN THE GUT. ITS FUNCTION
CC       IN THE BRAIN IS NOT CLEAR.
CC   -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
DR   PIR; A43001; A43001.
DR   PIR; PQ0012; PQ0012.
DR   INTERPRO; IPR001651; -.
DR   PROSITE; PS00259; GASTRIN; 1.
KW   Amidation; Sulfatation; Hormone.
FT   MOD_RES       2      2       SULFATATION.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  1064 MW;  DDCAA68378768B5A CRC64;
```

```
  Query Match             26.7%;  Score 12;  DB 1;  Length 8;
  Best Local Similarity   33.3%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        1 YLSGAD 6
            |:   |
Db        2 YMGWMD 7


RESULT 13
FARD_CALVO
ID   FARD_CALVO     STANDARD;      PRT;     9 AA.
AC   P41868;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 13.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; D44787; D44787.
KW   Neuropeptide; Amidation.
FT   MOD_RES      9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1028 MW;  22D10699C87AB6D8 CRC64;


  Query Match             26.7%;  Score 12;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        3 SGAD 6
             :| |
Db        1 AGQD 4


RESULT 14
OXYT_RABIT
ID   OXYT_RABIT     STANDARD;      PRT;     9 AA.
AC   P32878; P01188;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   OXYTOCIN (OCYTOCIN).
OS   Oryctolagus cuniculus (Rabbit), Hippopotamus amphibius (Hippopotamus),
OS   Balaenoptera physalus (Finback whale) (Common rorqual),
OS   Tachyglossus aculeatus aculeatus (Australian echidna), and
OS   Hydrolagus colliei (Spotted ratfish) (Pacific ratfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=RABBIT;
RX   MEDLINE; 72215060.
RA   Chauvet J., Chauvet M.-T., Acher R.;
RT   "Evolution of neurohypophyseal hormones: isolation of active
RT   principles from rabbits and rats.";
RL   Biochimie 53:1099-1104(1971).
RN   [2]
RP   SEQUENCE.
RC   SPECIES=H.AMPHIBIUS;
RX   MEDLINE; 71232719.
RA   Ferguson D.R., Pickering B.T.;
RT   "Arginine and lysine vasopressins in the hippopotamus
RT   neurohypophysis.";
RL   Gen. Comp. Endocrinol. 13:425-429(1969).
RN   [3]
RP   SEQUENCE.
RC   SPECIES=B.PHYSALUS;
RA   Acher R., Chauvet J., Chauvet M.-T.;
RT   "Isolation of finback whale oxytocin and vasopressin.";
RL   Nature 201:191-192(1964).
RN   [4]
RP   SEQUENCE.
RC   SPECIES=A.ACULEATUS;
RX   MEDLINE; 73223515.
RA   Acher R., Chauvet J., Chauvet M.-T.;
RT   "Neurohypophysial hormones and evolution of tetrapods.";
RL   Nature New Biol. 244:124-126(1973).
RN   [5]
RP   SEQUENCE.
RC   SPECIES=H.COLLIEI;
RX   MEDLINE; 70088110.
RA   Pickering B.T., Heller H.;
RT   "Oxytocin as a neurophyophysial hormone in the holocephalian
RT   elasmobranch fish, Hydrolagus colliei.";
RL   J. Endocrinol. 45:597-606(1969).
CC   -!- FUNCTION: OXYTOCIN CAUSES CONTRACTION OF THE SMOOTH MUSCLE OF THE
CC       UTERUS AND OF THE MAMMARY GLAND.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   PIR; A91466; A91466.
DR   PIR; A92774; A92774.
DR   PIR; A93147; A93147.
DR   PIR; A93408; A93408.
DR   PIR; B90667; B90667.
DR   PDB; 1XY1; 15-OCT-90.
DR   PDB; 1XY2; 15-OCT-90.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Hypothalamus; Amidation; 3D-structure.
FT   DISULFID     1      6
FT   MOD_RES      9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1010 MW;  17F8376EB456D04B CRC64;


  Query Match             26.7%;  Score 12;  DB 1;  Length 9;
  Best Local Similarity   28.6%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        1 YLSGADL 7
            |:    |
Db        2 YIQNCPL 8


RESULT 15
PGLR_DIAAB
ID   PGLR_DIAAB     STANDARD;      PRT;     9 AA.
AC   P81179;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ENDO-POLYGALACTURONASE (PG) (EC 3.2.1.15) (FRAGMENT).
OS   Diaprepes abbreviatus (Sugarcane rootstalk borer weevil).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC   Cucujiformia; Curculionidae; Entiminae; Entimini; Diaprepes.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LARVAL GUT;
RA   Doostdar H., McCollum T.G., Mayer R.T.;
RT   "Purification and characterization of an endo-polygalacturonase from
RT   the gut of West Indies sugarcane rootstalk borer weevil (Diaprepes
```

```
RT   abbreviatus L.) larvae.";
RL   Comp. Biochem. Physiol. 118B:861-867(1997).
CC   -!- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-ALPHA-D-
CC       GALACTOSIDURONIC LINKAGES IN PECTATE AND OTHER GALACTURONANS.
CC   -!- INDUCTION: INHIBITED BY CITRUS PGIP.
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS PROTEIN IS:
CC       9.4, ITS MW IS: 44.5 KDA.
CC   -!- SIMILARITY: WEAK TO OTHER POLYGALACTURONASES.
KW   Hydrolase; Glycosidase; Cell wall.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1041 MW;  1F49087042DB41BB CRC64;

  Query Match             26.7%;  Score 12;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |: |
Db       4 YVIG 7


RESULT  16
FAR3_HIRME
ID   FAR3_HIRME     STANDARD;      PRT;     4 AA.
AC   P42562;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE YLRF-AMIDE.
OS   Hirudo medicinalis (Medicinal leech).
OC   Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;
OC   Arynchobdellida; Hiruduniformes; Hirudinidae; Hirudo.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92195954.
RA   Evans B.D., Pohl J., Kartsonis M.A., Calabrese R.L.;
RT   "Identification of RFamide neuropeptides in the medicinal leech.";
RL   Peptides 12:897-908(1991).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       4      4       AMIDATION.
SQ   SEQUENCE   4 AA;  598 MW;  69D4073B30000000 CRC64;


  Query Match             24.4%;  Score 11;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YL 2
           ||
Db       1 YL 2


RESULT  17
PRCT_PERAM
ID   PRCT_PERAM     STANDARD;      PRT;     5 AA.
AC   P01373;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   PROCTOLIN.
OS   Periplaneta americana (American cockroach),
OS   Limulus polyphemus (Atlantic horseshoe crab), and
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blattoidea; Blattidae; Periplaneta.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=P.AMERICANA;
RX   MEDLINE; 76074708.
RA   Starratt A.N., Brown B.E.;
RT   "Structure of the pentapeptide proctolin, a proposed neurotransmitter
RT   in insects.";
RL   Life Sci. 17:1253-1256(1975).
RN   [2]
RP   BIOLOGICAL SOURCE.
RC   SPECIES=P.AMERICANA;
RX   MEDLINE; 81225865.
RA   O'Shea M., Adams M.E.;
RT   "Pentapeptide (proctolin) associated with an identified neuron.";
RL   Science 213:567-569(1981).
RN   [3]
RP   SEQUENCE.
RC   SPECIES=L.POLYPHEMUS;
RX   MEDLINE; 90287800.
RA   Groome J.R., Tillinghast E.K., Townley M.A., Vetrovs A.,
RA   Watson W.H. III, Hunt D.F., Griffin P.R., Alexander J.E.,
RA   Shabanowitz J.;
RT   "Identification of proctolin in the central nervous system of the
RT   horseshoe crab, Limulus polyphemus.";
RL   Peptides 11:205-211(1990).
RN   [4]
RP   SEQUENCE.
RC   SPECIES=C.MAENAS;
RX   MEDLINE; 86232789.
RA   Stangier J., Dircksen H., Keller R.;
RT   "Identification and immunocytochemical localization of proctolin in
RT   pericardial organs of the shore crab, Carcinus maenas.";
RL   Peptides 7:67-72(1986).
CC   -!- FUNCTION: STIMULATES CARDIAC OUTPUT AND HINDGUT MOTILITY,
CC       MODULATES VISCERAL AND SKELETAL MUSCLE IN MANY ARTHROPODS.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE LATERAL WHITE NEURONS AND IN
CC       THE CRAB PERICARDIAL ORGANS.
DR   PIR; A01644; HOROHA.
DR   PIR; A60411; A60411.
KW   Neuropeptide.
SQ   SEQUENCE   5 AA;  649 MW;  71B7673B44600000 CRC64;


  Query Match             24.4%;  Score 11;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YL 2
           ||
Db       2 YL 3


RESULT  18
CIP2_MYTED
ID   CIP2_MYTED     STANDARD;      PRT;     6 AA.
AC   P13737;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   CONTRACTION-INHIBITING PEPTIDE II (MIP II).
OS   Mytilus edulis (Blue mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytiloidea; Mytilidae; Mytilus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PEDAL GANGLION;
RX   MEDLINE; 88240357.
RA   Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.;
RT   "Structures and actions of Mytilus inhibitory peptides.";
RL   Biochem. Biophys. Res. Commun. 152:1376-1382(1988).
CC   -!- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN
CC       MUSCLES.
CC   -!- SIMILARITY: TO MIP I.
DR   PIR; B27696; B27696.
KW   Hormone; Amidation.
FT   MOD_RES       6      6       AMIDATION.
SQ   SEQUENCE   6 AA;  621 MW;  72C9C6876DD81000 CRC64;
```

```
  Query Match             24.4%;  Score 11;  DB 1;  Length 6;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GADL 7
           || :
Db       1 GAPM 4


RESULT  19
FAR2_ASCSU
ID   FAR2_ASCSU     STANDARD;      PRT;     7 AA.
AC   P31890;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE AF2.
OS   Ascaris suum (Pig roundworm) (Ascaris lumbricoides), and
OS   Panagrellus redivivus.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC   Ascarididae; Ascaris.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=A.SUUM;
RX   MEDLINE; 93324431.
RA   Cowden C., Stretton A.O.W.;
RT   "AF2, an Ascaris neuropeptide: isolation, sequence, and bioactivity.";
RL   Peptides 14:423-430(1993).
RN   [2]
RP   SEQUENCE.
RC   SPECIES=P.REDIVIVUS;
RX   MEDLINE; 95060998.
RA   Maule A.G., Shaw C., Bowman J.W.;
RT   "The FMRFamide-like neuropeptide AF2 (Ascaris suum) is present in the
RT   free-living nematode, Panagrellus redivivus (Nematoda, Rhabditida).";
RL   Parasitology 109:351-356(1994).
CC   -!- FUNCTION: HAS EFFECTS ON MUSCLE TENSION.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE NERVE CORDS AND A VARIETY OF
CC       GANGLIA PARTICULARLY IN THE ANTERIOR REGIONS.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  992 MW;  69D4073B5B11E350 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YL 2
           ||
Db       4 YL 5


RESULT  20
GFRP_MOUSE
ID   GFRP_MOUSE     STANDARD;      PRT;     7 AA.
AC   P99025;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GTP CYCLOHYDROLASE I FEEDBACK REGULATORY PROTEIN (P35) (FRAGMENT).
GN   GFRP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LIVER;
RA   Sanchez J.-C., Rouge V., Frutiger S., Hughes G., Yan J.X.,
RA   Hoogland C., Appel R.D., Binz P.-A., Hochstrasser D.F.,
RA   Cowthorne M.;
RL   Submitted (AUG-1998) to the SWISS-PROT data bank.
CC   -!- FUNCTION: MEDIATES TETRAHYDROBIOPTERIN INHIBITION OF GTP
CC       CYCLOHYDROLASE I. THIS INHIBITION IS REVERSED BY L-PHENYLALANINE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
DR   SWISS-2DPAGE; P99025; MOUSE.
FT   INIT_MET      0      0
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  806 MW;  71B5B057273B4700 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YL 2
           ||
Db       2 YL 3


RESULT  21
LANC_CARUI
ID   LANC_CARUI     STANDARD;      PRT;     7 AA.
AC   P36960;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   LANTIBIOTIC CARNOCIN UI49 (FRAGMENT).
OS   Carnobacterium sp. (strain UI49).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC   Carnobacterium.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92321768.
RA   Stoffels G., Nissen-Meyer J., Gudmundsdottir A., Sletten K., Holo H.,
RA   Nes I.F.;
RT   "Purification and characterization of a new bacteriocin isolated from
RT   a Carnobacterium sp.";
RL   Appl. Environ. Microbiol. 58:1417-1422(1992).
CC   -!- FUNCTION: LANTHIONINE-CONTAINING PEPTIDE ANTIBIOTIC (LANTIBIOTIC).
CC       ACTIVE ON GRAM-POSITIVE BACTERIA.
KW   Antibiotic; Lantibiotic.
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  786 MW;  741776D05B05B810 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 7;
  Best Local Similarity   25.0%;  Pred. No. 8.8e+04;
  Matches    1;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GADL 7
           |:::
Db       1 GSEI 4


RESULT  22
UF03_MOUSE
ID   UF03_MOUSE     STANDARD;      PRT;     7 AA.
AC   P38641;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE OF FIBROBLASTS (P36) (FRAGMENT).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=FIBROBLAST;
RX   MEDLINE; 95009907.
RA   Merrick B.A., Patterson R.M., Wichter L.L., He C., Selkirk J.K.;
```

```
RT   "Separation and sequencing of familiar and novel murine proteins
RT   using preparative two-dimensional gel electrophoresis.";
RL   Electrophoresis 15:735-745(1994).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 5.1, ITS MW IS: 36 KDA.
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  842 MW;  6AA72B1DDB1B1180 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ADLN 8
           |:|:
Db       4 AELD 7


RESULT  23
CAD1_ENTFA
ID   CAD1_ENTFA     STANDARD;      PRT;     8 AA.
AC   P13268;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   SEX PHEROMONE CAD1.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 85051889.
RA   Mori M., Sagakami Y., Narita M., Isogai A., Fujino M., Kitada C.,
RA   Craig R.A., Clewell D.B., Suzuki A.;
RT   "Isolation and structure of the bacterial sex pheromone, cAD1, that
RT   induces plasmid transfer in Streptococcus faecalis.";
RL   FEBS Lett. 178:97-100(1984).
CC   -!- FUNCTION: CAD1 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE
CC       HEMOLYSIN PLASMID PAD1.
KW   Pheromone.
SQ   SEQUENCE   8 AA;  819 MW;  047DD732C735B9C7 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           |:|
Db       6 LAG 8


RESULT  24
FAR3_HOMAM
ID   FAR3_HOMAM     STANDARD;      PRT;     8 AA.
AC   P41486;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE 3 (FLI 3) (F2).
OS   Homarus americanus (American lobster).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC   Nephropoidea; Nephropidae; Homarus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PERICARDIAL ORGANS;
RX   MEDLINE; 88116164.
RA   Trimmer B.A., Kobierski L.A., Kravitz E.A.;
RT   "Purification and characterization of FMRFamidelike immunoreactive
RT   substances from the lobster nervous system: isolation and sequence
RT   analysis of two closely related peptides.";
RL   J. Comp. Neurol. 266:16-26(1987).
```

```
CC   -!- MISCELLANEOUS: PERICARDIAL ORGANS RELEASE THIS PEPTIDE WITH 100 MM
CC       POTASSIUM IN THE PRESENCE OF CALCIUM.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  1054 MW;  C6D40729C4540AB5 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 ADLN 8
            :| |
Db       1 SDRN 4


RESULT  25
FAR8_CALVO
ID   FAR8_CALVO     STANDARD;      PRT;     8 AA.
AC   P41863;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 8.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; H41978; H41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  957 MW;  72D40699CAA44DD8 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GAD 6
           ||:
Db       1 GAN 3


RESULT  26
D1_NEPNO
ID   D1_NEPNO       STANDARD;      PRT;     9 AA.
AC   P24816;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   01-MAR-1992 (Rel. 21, Last annotation update)
DE   GASTRIN/CHOLECYSTOKININ-LIKE PEPTIDE D1.
OS   Nephrops norvegicus (Norway lobster).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC   Nephropoidea; Nephropidae; Nephrops.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=STOMACH;
RX   MEDLINE; 92082847.
```

```
RA   Favrel P., Kegel G., Sedlmeier D., Keller R., van Wormhoudt A.;
RT   "Structure and biological activity of crustacean gastrointestinal
RT   peptides identified with antibodies to gastrin/cholecystokinin.";
RL   Biochimie 73:1233-1239(1991).
CC   -!- FUNCTION: MAY CONTROL DIGESTION PROCESSES IN CRUSTACEANS.
CC   -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
DR   PIR; A48398; A48398.
KW   Hormone.
SQ   SEQUENCE   9 AA;  1038 MW;  60EC79CAB6D8787B CRC64;


  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 GAD 6
            | |
Db        4 GQD 6.


RESULT  27
FAR5_CALVO
ID   FAR5_CALVO     STANDARD;      PRT;     9 AA.
AC   P41860;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 5.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; E41978; E41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1068 MW;  39D10699CAB6D867 CRC64;


  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 GAD 6
            | |
Db        3 GQD 5


RESULT  28
FAR7_CALVO
ID   FAR7_CALVO     STANDARD;      PRT;     9 AA.
AC   P41862;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 7.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; G41978; G41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1081 MW;  E1D10699CAB6D86A CRC64;


  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 GAD 6
            | |
Db        3 GQD 5


RESULT  29
MOSF_CLYJA
ID   MOSF_CLYJA     STANDARD;      PRT;     9 AA.
AC   P19853;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   [PHE-6]-MOSACT.
OS   Clypeaster japonicus (Sand dollar).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Gnathostomata; Clypeasteroida;
OC   Clypeasteridae; Clypeaster.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=EGG JELLY;
RA   Suzuki N., Kurita M., Yoshino K.I., Kajiura H., Nomura K.,
RA   Yamaguchi M.;
RT   "Purification and structure of mosact and its derivatives from the
RT   egg jelly of the sea urchin Clypeaster japonicus.";
RL   Zool. Sci. 4:649-656(1987).
CC   -!- FUNCTION: STIMULATES SPERM RESPIRATION AND MOTILITY.
DR   PIR; JN0027; JN0027.
SQ   SEQUENCE   9 AA;  924 MW;  93245729CDC5BAB5 CRC64;


  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSG 4
            :| |
Db        6 FLIG 9


RESULT  30
ACH1_ACHFU
ID   ACH1_ACHFU     STANDARD;      PRT;     4 AA.
AC   P35904;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   ACHATIN-I.
OS   Achatina fulica (Giant African snail).
OC   Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC   Achatinacea; Achatinidae; Achatina.
RN   [1]
RP   SEQUENCE, CHARACTERIZATION, AND SYNTHESIS.
RC   STRAIN=FERUSSAC; TISSUE=GANGLION;
```

```
RX   MEDLINE; 89273551.
RA   Kamatani Y., Minakata H., Kenny P.T.M., Iwashita T., Watanabe K.,
RA   Funase K., Sun X.P., Yongsiri A., Kim K.H., Novales-Li P.,
RA   Novales E.T., Kanapi C.G., Takeuchi H., Nomoto K.;
RT   "Achatin-I, an endogenous neuroexcitatory tetrapeptide from Achatina
RT   fulica Ferussac containing a D-amino acid residue.";
RL   Biochem. Biophys. Res. Commun. 160:1015-1020(1989).
RN   [2]
RP   CHARACTERIZATION.
RC   STRAIN=FERUSSAC; TISSUE=HEART ATRIUM;
RX   MEDLINE; 91264856.
RA   Fujimoto K., Kubota I., Yasuda-Kamatani Y., Minakata H., Nomoto K.,
RA   Yoshida M., Harada A., Muneoka Y., Kobayashi M.;
RT   "Purification of achatin-I from the atria of the African giant snail,
RT   Achatina fulica, and its possible function.";
RL   Biochem. Biophys. Res. Commun. 177:847-853(1991).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY.
RX   MEDLINE; 93014529.
RA   Ishida T., In Y., Doi M., Inoue M., Yasuda-Kamatani Y., Minakata H.,
RA   Iwashita T., Nomoto K.;
RT   "Crystal structure and molecular conformation of achatin-I
RT   (H-Gly-D-Phe-Ala-Asp-OH), an endogenous neuropeptide containing a
RT   D-amino acid residue.";
RL   Int. J. Pept. Protein Res. 39:258-264(1992).
CC   -!- FUNCTION: NEUROEXCITATORY PEPTIDE; INCREASES THE IMPULSE FREQUENCY
CC       AND PRODUCES A SPIKE BROADENING OF THE IDENTIFIED HEART EXCITATORY
CC       NEURON (PON); ALSO ENHANCES THE AMPLITUDE AND FREQUENCY OF THE
CC       HEART BEAT. HAS ALSO AN EFFECT ON SEVERAL OTHER MUSCLES.
DR   PIR; A32480; A32480.
KW   Hormone; D-amino acid.
FT   MOD_RES       2      2       D-PHENYLALANINE.
SQ   SEQUENCE   4 AA;  408 MW;  6AADD9C810000000 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 AD 6
           ||
Db       3 AD 4


RESULT 31
UXA4_CHLTR
ID   UXA4_CHLTR     STANDARD;      PRT;     5 AA.
AC   P38005;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE FROM ELEMENTARY BODY (FRAGMENT).
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE.
RC   STRAIN=L2/434/BU;
RA   Bini L., Santucci A., Magi B., Marzocchi B., Sanchez-Campillo M.,
RA   Comanducci M., Christianen G., Birkelund S., Vtretou E., Ratti G.,
RA   Pallini V.;
RL   Submitted (SEP-1994) to the SWISS-PROT data bank.
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 4.5, ITS MW IS: 28 KDA.
FT   NON_TER       5      5
SQ   SEQUENCE   5 AA;  474 MW;  75BAA865AA800000 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SG 4
           ||
```

```
Db       2 SG 3


RESULT 32
TMOF_SARBU
ID   TMOF_SARBU     STANDARD;      PRT;     6 AA.
AC   P41495;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   TRYPSIN-MODULATING OOSTATIC FACTOR (TMOF).
OS   Sarcophaga bullata (Grey flesh fly) (Neobellieria bullata).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Sarcophagidae; Sarcophaga.
RN   [1]
RP   SEQUENCE, AND SYNTHESIS.
RC   TISSUE=OVARY;
RX   MEDLINE; 94211930.
RA   Bylemans D., Borovsky D., Hunt D.F., Shabanowitz J., Grauwels L.,
RA   de Loof A.;
RT   "Sequencing and characterization of trypsin modulating oostatic
RT   factor (TMOF) from the ovaries of the grey fleshfly, Neobellieria
RT   (Sarcophaga) bullata.";
RL   Regul. Pept. 50:61-72(1994).
CC   -!- FUNCTION: HAS AN OOSTATIC ACTIVITY. INHIBITS TRYPSIN BIOSYNTHESIS
CC       IN THE MIDGUT WHICH INDIRECTLY REDUCES THE VITELLOGENIN
CC       CONCENTRATION IN THE HEMOLYMPH RESULTING IN INHIBITION OF OOCYTE
CC       DEVELOPMENT.
CC   -!- DEVELOPMENTAL STAGE: SYNTHESIZED AND RELEASED FROM FOLLICULAR
CC       EPITHELIUM AFTER A BLOOD MEAL.
KW   Hormone.
SQ   SEQUENCE   6 AA;  695 MW;  61E72451B7642000 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 NL 9
           ||
Db       4 NL 5


RESULT 33
TRPI_PSEPU
ID   TRPI_PSEPU     STANDARD;      PRT;     6 AA.
AC   P36414;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   TRPBA OPERON TRANSCRIPTIONAL ACTIVATOR (FRAGMENT).
GN   TRPI.
OS   Pseudomonas putida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PPG1 C1S;
RX   MEDLINE; 89335826.
RA   Eberly L., Crawford I.P.;
RT   "DNA sequence of the tryptophan synthase genes of Pseudomonas
RT   putida.";
RL   Biochimie 71:521-531(1989).
CC   -!- FUNCTION: ACTIVATES THE EXPRESSION OF THE TRPBA GENES ENCODING
CC       THE TWO TRYPTOPHAN SYNTHASE SUBUNITS. IN THE ABSENCE OF THE
CC       INDUCER (INDOLEGLYCEROL PHOSPHATE), TRPI BINDS UPSTREAM OF THE
CC       TRPAB OPERON, OVERLAPPING ITS OWN PROMOTER REGION.
CC   -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC       REGULATORS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X13299; CAA31660.1; -.
DR   INTERPRO; IPR000847; -.
DR   PROSITE; PS00044; HTH_LYSR_FAMILY; PARTIAL.
KW   Tryptophan biosynthesis; Transcription regulation; Activator;
KW   DNA-binding.
FT   NON_TER       6      6
SQ   SEQUENCE   6 AA;  683 MW;  77672AA1EDD6F000 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DL 7
           ||
Db       4 DL 5


RESULT  34
ALL2_CARMA
ID   ALL2_CARMA     STANDARD;      PRT;     7 AA.
AC   P81805;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 2.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   MOD_RES       7      7       AMIDATION (POTENTIAL).
SQ   SEQUENCE   7 AA;  770 MW;  672879CDCB5DDB70 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |  |
Db       3 YAFG 6


RESULT  35
ALL3_CARMA
ID   ALL3_CARMA     STANDARD;      PRT;     7 AA.
AC   P81806;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 3.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Multigene family.
SQ   SEQUENCE   7 AA;  796 MW;  672879CDCB476B70 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |  |
Db       3 YAFG 6


Search completed: December 16, 2000, 04:23:31
Job time: 4566 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 03:35:12 ; Search time 89.11 Seconds
                                                (without alignments)
                                                6.409 Million cell updates/sec

Title:           US-09-529-121-2
Perfect score:   45
Sequence:        1 YLSGADLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters:      793

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       PIR_65:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                               SUMMARIES
               %
Result        Query
  No.   Score Match Length DB  ID                   Description
--------------------------------------------------------------------------
    1     16   35.6      7  2  PC2370               probable H+-transp
    2     16   35.6      7  2  PT0542               T-cell receptor be
    3     16   35.6      8  2  JS0318               leucokinin VIII -
    4     16   35.6      9  1  AKLQIM               locustamyoinhibiti
    5     16   35.6      9  2  A57444               neuropeptide Grb-A
    6     16   35.6      9  2  B57444               neuropeptide Grb-A
    7     16   35.6      9  4  I57650               hemoglobin alpha c
    8     15   33.3      5  2  PT0540               T-cell receptor be
    9     15   33.3      6  2  PT0726               T-cell receptor be
   10     15   33.3      7  2  PT0526               T-cell receptor be
   11     15   33.3      7  2  PT0676               T-cell receptor be
   12     15   33.3      8  2  T13818               cytochrome oxidase
   13     15   33.3      9  2  F41978               calliFMRFamide 6 -
   14     15   33.3      9  2  PT0288               Ig heavy chain CRD
   15     15   33.3      9  2  G41946               T-cell receptor ga
   16     14   31.1      4  2  A26209               protein-glutamine
   17     14   31.1      5  2  PT0679               T-cell receptor be
   18     14   31.1      6  2  PT0605               T-cell receptor be
   19     14   31.1      6  2  PT0593               T-cell receptor be
   20     14   31.1      7  2  S20446               elastase - Pseudom
   21     14   31.1      7  2  PT0654               T-cell receptor be
   22     14   31.1      7  2  PT0722               T-cell receptor be
   23     14   31.1      8  2  A21440               variant surface gl
   24     14   31.1      8  2  A41117               acetylcholinestera
   25     14   31.1      8  2  PN0043               phosphatidylethano
   26     14   31.1      8  2  PT0557               T-cell receptor be
   27     14   31.1      9  2  A61364               isotocin - common
   28     14   31.1      9  2  C57444               neuropeptide Grb-A
   29     13   28.9      4  2  S43959               Ig mu chain V regi
```

```
   30     13   28.9      7  2  I50210               gene c-rel protein
   31     13   28.9      8  2  PT0547               T-cell receptor be
   32     13   28.9      9  2  A44873               caldesmon - rabbit
   33     13   28.9      9  2  QDRB                 delta sleep-induci
   34     13   28.9      9  2  C41170               photosystem II pro
   35     13   28.9      9  2  PH0935               T-cell receptor be
   36     13   28.9      9  2  PH0918               T-cell receptor be
   37     12   26.7      6  2  I51434               H4 histone - Afric
   38     12   26.7      7  2  S16364               opacity protein P.
   39     12   26.7      7  2  B35890               RNA-directed DNA p
   40     12   26.7      7  2  PN0649               alpha-dextrin endo
   41     12   26.7      7  2  S29735               polyphosphate--glu
   42     12   26.7      8  2  PQ0012               cholecystokinin -
   43     12   26.7      8  2  A43001               cholecystokinin -
   44     12   26.7      8  2  PQ0701               unidentified 6.5/3
   45     12   26.7      8  2  PL0184               capsid protein VP-
   46     12   26.7      8  2  S65647               2-hydroxyglutaryl-
   47     12   26.7      8  2  G33098               205K exoantigen -
   48     12   26.7      8  2  S21663               neuropeptide - flo
   49     12   26.7      8  2  PT0522               T-cell receptor be
   50     12   26.7      9  2  A91466               oxytocin - hippopo
   51     12   26.7      9  2  A92774               oxytocin - spotted
   52     12   26.7      9  2  A93147               oxytocin - finback
   53     12   26.7      9  2  A93408               oxytocin - Austral
   54     12   26.7      9  2  B90667               oxytocin - rabbit
   55     12   26.7      9  2  D44787               calliFMRFamide 13
   56     12   26.7      9  2  PT0268               Ig heavy chain CRD
   57     11   24.4      5  1  HOROHA               proctolin - Americ
   58     11   24.4      5  2  A41225               copper resistance
   59     11   24.4      5  2  B31836               20K protein - Rick
   60     11   24.4      5  2  A60411               proctolin - Atlant
   61     11   24.4      5  2  PT0267               Ig heavy chain CRD
   62     11   24.4      5  2  C23751               spinal cord peptid
   63     11   24.4      5  2  PT0651               T-cell receptor be
   64     11   24.4      6  2  B44510               hypothetical prote
   65     11   24.4      6  2  B27696               contraction-inhibi
   66     11   24.4      6  2  PT0280               Ig heavy chain CRD
   67     11   24.4      6  2  B35640               cerebellar degener
   68     11   24.4      7  2  S71867               glutathione transf
   69     11   24.4      7  2  S42407               gramicidin S synth
   70     11   24.4      7  2  S09066               globulin IV alpha
   71     11   24.4      7  2  S78024               ribosomal protein
   72     11   24.4      7  2  S58797               serine/threonine-s
   73     11   24.4      7  2  B48394               major fat-globule
   74     11   24.4      7  2  A58718               carnocin UI49 - Ca
   75     11   24.4      8  2  PC4131               hypothetical prote

                              ALIGNMENTS

RESULT   1
PC2370
probable H+-transporting ATP synthase (EC 3.6.1.34) alpha chain [similarity] - Bacill
N;Alternate names: unidentified 78K protein
C;Species: Bacillus cereus
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: PC2370
R;Matsuno, K.; Miyamoto, T.; Yamaguchi, K.; Sayed, M.A.; Kajiwara, T.; Hatano, S.
Biosci. Biotechnol. Biochem. 59, 231-235, 1995
A;Title: Identification of DNA-binding proteins changed after induction of sporulatio
A;Reference number: PC2369; MUID:95218265
A;Accession: PC2370
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-7 <MAS>
C;Keywords: ATP biosynthesis; hydrolase


  Query Match             35.6%;  Score 16;  DB 2;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       6 DLN 8
           |||
Db       2 DLN 4


RESULT   2
PT0542
T-cell receptor beta chain V-D-J region (126-1BA) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0542
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0542
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-7 <FEE>
A;Experimental source: day 18 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             35.6%;  Score 16;  DB 2;  Length 7;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAD 6
           || |
Db       2 SGGD 5


RESULT   3
JS0318
leucokinin VIII - Madeira cockroach
C;Species: Leucophaea maderae (Madeira cockroach)
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 20-Jun-2000
C;Accession: JS0318
R;Holman, G.M.; Cook, B.J.; Nachman, R.J.
Comp. Biochem. Physiol. C 88, 31-34, 1987
A;Title: Isolation, primary structure and synthesis of leucokinins VII and VIII: the fir
A;Reference number: JS0317
A;Accession: JS0318
A;Molecule type: protein
A;Residues: 1-8 <HOL>
C;Comment: Leucokinins, a family of cephalomyotropic peptides, stimulate contractile act
C;Keywords: amidated carboxyl end; cephalomyotropic peptide
F;8/Modified site: amidated carboxyl end (Gly) #status experimental


  Query Match             35.6%;  Score 16;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GAD 6
           |||
Db       1 GAD 3


RESULT   4
AKLQIM
locustamyoinhibiting peptide - migratory locust
C;Species: Locusta migratoria (migratory locust)
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 20-Mar-1998
C;Accession: A60065
R;Schoofs, L.; Holman, G.M.; Hayes, T.K.; Nachman, R.J.; De Loof, A.
Regul. Pept. 36, 111-119, 1991
A;Title: Isolation, identification and synthesis of locustamyoinhibiting peptide (LOM-MI
A;Reference number: A60065; MUID:92179466
A;Accession: A60065
A;Molecule type: protein
A;Residues: 1-9 <SCH>
C;Comment: This peptide hormone suppresses spontaneous contractions of the hindgut and d
```

```
C;Superfamily: locustamyoinhibiting peptide
C;Keywords: amidated carboxyl end; hormone
F;9/Modified site: amidated carboxyl end (Trp) #status experimental


  Query Match             35.6%;  Score 16;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DLN 8
           |||
Db       4 DLN 6


RESULT   5
A57444
neuropeptide Grb-AST B1 - two-spotted cricket
C;Species: Gryllus bimaculatus (two-spotted cricket)
C;Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 26-Jan-1996
C;Accession: A57444
R;Lorenz, M.W.; Kellner, R.; Hoffmann, K.H.
J. Biol. Chem. 270, 21103-21108, 1995
A;Title: A family of neuropeptides that inhibit juvenile hormone biosynthesis in the
A;Reference number: A57444; MUID:95403341
A;Accession: A57444
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <LOR>


  Query Match             35.6%;  Score 16;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DLN 8
           |||
Db       4 DLN 6


RESULT   6
B57444
neuropeptide Grb-AST B2 - two-spotted cricket
C;Species: Gryllus bimaculatus (two-spotted cricket)
C;Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 26-Jan-1996
C;Accession: B57444
R;Lorenz, M.W.; Kellner, R.; Hoffmann, K.H.
J. Biol. Chem. 270, 21103-21108, 1995
A;Title: A family of neuropeptides that inhibit juvenile hormone biosynthesis in the
A;Reference number: A57444; MUID:95403341
A;Accession: B57444
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <LOR>


  Query Match             35.6%;  Score 16;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DLN 8
           |||
Db       4 DLN 6


RESULT   7
I57650
hemoglobin alpha chain - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 31-Jul-1997 #text_change 20-Apr-2000
C;Accession: I57650
R;Whitelaw, E.; Hogben, P.; Hanscombe, O.; Proudfoot, N.J.
Mol. Cell. Biol. 9, 241-251, 1989
A;Title: Transcriptional promiscuity of the human alpha-globin gene.
```

```
A;Reference number: I57650; MUID:89181576
A;Accession: I57650
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-9 <WHI>
A;Cross-references: GB:M23454; NID:g340922; PIDN:AAA52629.1; PID:g553329
A;Note: engineered sequence; this sequence was not determined in this report
C;Genetics:
A;Gene: GDB:HBA1
A;Cross-references: GDB:119293
A;Map position: 16p13.3-16p13.3

  Query Match             35.6%;  Score 16;  DB 4;  Length 9;
  Best Local Similarity   80.0%;  Pred. No. 1.8e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LSGAD 6
           || ||
Db       3 LSPAD 7


RESULT   8
PT0540
T-cell receptor beta chain V-D-J region (126-1L) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0540
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0540
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-5 <FEE>
A;Experimental source: day 18 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             33.3%;  Score 15;  DB 2;  Length 5;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAD 6
           || |
Db       2 SGED 5


RESULT   9
PT0726
T-cell receptor beta chain V-D-J region (161-2D) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0726
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0726
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-6 <FEE>
A;Experimental source: newborn thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             33.3%;  Score 15;  DB 2;  Length 6;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAD 6
           || |
```

```
Db       2 SGED 5


RESULT  10
PT0526
T-cell receptor beta chain V-D-J region (100-4E) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0526
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0526
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-7 <FEE>
A;Experimental source: adult thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             33.3%;  Score 15;  DB 2;  Length 7;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAD 6
           || |
Db       2 SGED 5


RESULT  11
PT0676
T-cell receptor beta chain V-D-J region (140-1AL) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0676
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0676
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-7 <FEE>
A;Experimental source: day 18 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             33.3%;  Score 15;  DB 2;  Length 7;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAD 6
           || |
Db       2 SGED 5


RESULT  12
T13818
cytochrome oxidase subunit I - Atlantic hagfish mitochondrion (fragment)
C;Species: mitochondrion Myxine glutinosa (Atlantic hagfish)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C;Accession: T13818
R;Delarbre, C.; Barriel, V.; Tillier, S.; Janvier, P.; Gachelin, G.
Mol. Biol. Evol. 14, 807-813, 1997
A;Title: The main features of the craniate mitochondrial DNA between the ND1 and the
A;Reference number: Z17775; MUID:97398704
A;Accession: T13818
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-8 <DEL>
A;Cross-references: EMBL:Y09527; NID:e1011465; PID:e329906; PIDN:CAA70718.1
C;Genetics:
```

```
A;Genome: mitochondrion
A;Note: COI
C;Keywords: mitochondrion


  Query Match            33.3%;  Score 15;  DB 2;  Length 8;
  Best Local Similarity  100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLS 3
           |||
Db       2 YLS 4

RESULT  13
F41978
calliFMRFamide 6 - bluebottle fly (Calliphora vomitoria)
C;Species: Calliphora vomitoria
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C;Accession: F41978
R;Duve, H.; Johnsen, A.H.; Sewell, J.C.; Scott, A.G.; Orchard, I.; Rehfeld, J.F.; Thorpe
Proc. Natl. Acad. Sci. U.S.A. 89, 2326-2330, 1992
A;Title: Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH-2 neuropeptides (desi
A;Reference number: A41978; MUID:92196111
A;Accession: F41978
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <DUV>
C;Keywords: amidated carboxyl end; neuropeptide
F;9/Modified site: amidated carboxyl end (Phe) #status experimental


  Query Match            33.3%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity  75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAD 6
             || |
Db       2 SGQD 5

RESULT  14
PT0288
Ig heavy chain CRD3 region (clone 4-106) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C;Accession: PT0288
R;Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991
A;Title: Preferential utilization of specific immunoglobulin heavy chain diversity and j
A;Reference number: PT0222; MUID:91108337
A;Accession: PT0288
A;Molecule type: DNA
A;Residues: 1-9 <YAM>
A;Experimental source: B lymphocyte
C;Keywords: heterotetramer; immunoglobulin


  Query Match            33.3%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity  75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           | ||
Db       5 YSSG 8

RESULT  15
G41946
T-cell receptor gamma chain (2t.23) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 07-May-1999
C;Accession: G41946
```

```
R;Whetsell, M.; Mosley, R.L.; Whetsell, L.; Schaefer, F.V.; Miller, K.S.; Klein, J.R.
Mol. Cell. Biol. 11, 5902-5909, 1991
A;Title: Rearrangement and junctional-site sequence analyses of T-cell receptor gamma
A;Reference number: A41946; MUID:92049316
A;Accession: G41946
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-9 <WHE>
C;Keywords: T-cell receptor


  Query Match            33.3%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity  75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           | ||
Db       5 YSSG 8

RESULT  16
A26209
protein-glutamine gamma-glutamyltransferase (EC 2.3.2.13) - guinea pig (fragment)
C;Species: Cavia porcellus (guinea pig)
C;Date: 10-Sep-1987 #sequence_revision 10-Sep-1987 #text_change 03-Mar-1995
C;Accession: A26209
R;Connellan, J.M.; Chung, S.I.; Whetzel, N.K.; Bradley, L.M.; Folk, J.E.
J. Biol. Chem. 246, 1093-1098, 1971
A;Title: Structural properties of guinea pig liver transglutaminase.
A;Reference number: A26209; MUID:71111415
A;Accession: A26209
A;Molecule type: protein
A;Residues: 1-4 <CON>
A;Experimental source: liver
C;Keywords: aminoacyltransferase


  Query Match            31.1%;  Score 14;  DB 2;  Length 4;
  Best Local Similarity  100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ADL 7
           |||
Db       2 ADL 4

RESULT  17
PT0679
T-cell receptor beta chain V-D-J region - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0679; PT0708
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0679
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-5 <FEE>
A;Experimental source: day 18 fetal thymus, strain BALB/c, 154-2J
A;Accession: PT0708
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-5 <FE2>
A;Experimental source: newborn thymus, strain BALB/c, 161-2B
C;Keywords: T-cell receptor


  Query Match            31.1%;  Score 14;  DB 2;  Length 5;
  Best Local Similarity  75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy       3 SGAD 6
           || |
Db       2 SGDD 5

RESULT  18
PT0605
T-cell receptor beta chain V-D-J region (120-1L) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0605
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0605
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-6 <FEE>
A;Experimental source: newborn thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGA 5
           |||
Db       2 SGA 4


RESULT  19
PT0593
T-cell receptor beta chain V-D-J region (159-1F) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0593
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0593
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-6 <FEE>
A;Experimental source: day 19 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGA 5
           |||
Db       4 SGA 6


RESULT  20
S20446
elastase - Pseudomonas aeruginosa
C;Species: Pseudomonas aeruginosa
C;Date: 19-Mar-1997 #sequence_revision 24-Jul-1997 #text_change 24-Jul-1997
C;Accession: S20446
R;Kessler, E.; Safrin, M.; Peretz, M.; Burstein, Y.
FEBS Lett. 299, 291-293, 1992
A;Title: Identification of cleavage sites involved in proteolytic processing of Pseudomo
A;Reference number: S20446; MUID:92183956
A;Accession: S20446
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-7 <KES>
```

```
  Query Match             31.1%;  Score 14;  DB 2;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ADL 7
           |||
Db       2 ADL 4


RESULT  21
PT0654
T-cell receptor beta chain V-D-J region (121-1BK) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0654
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0654
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-7 <FEE>
A;Experimental source: day 4 postnatal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGA 5
           |||
Db       2 SGA 4


RESULT  22
PT0722
T-cell receptor beta chain V-D-J region (135-1G) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0722
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0722
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-7 <FEE>
A;Experimental source: newborn thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 7;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAD 6
           || |
Db       2 SGDD 5


RESULT  23
A21440
variant surface glycoprotein pSLc1 - Trypanosoma brucei (fragment)
C;Species: Trypanosoma brucei
C;Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 20-Mar-1998
C;Accession: A21440
R;Parsons, M.; Nelson, R.G.; Watkins, K.P.; Agabian, N.
Cell 38, 309-316, 1984
```

```
A;Title: Trypanosome mRNAs share a common 5' spliced leader sequence.
A;Reference number: A90853; MUID:84282716
A;Accession: A21440
A;Molecule type: mRNA
A;Residues: 1-8 <PAR>
A;Cross-references: GB:K02195; NID:g162150; PID:g162151
C;Keywords: glycoprotein


  Query Match             31.1%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   33.3%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 LSGADL 7
              :||  ::
Db          1 MSGKEV 6


RESULT  24
A41117
acetylcholinesterase (EC 3.1.1.7), venom - Asian cobra (fragment)
C;Species: Naja naja oxiana (Asian cobra, Oxus cobra)
C;Date: 27-Mar-1992 #sequence_revision 27-Mar-1992 #text_change 23-Jun-1993
C;Accession: A41117
R;Kreienkamp, H.J.; Weise, C.; Raba, R.; Aaviksaar, A.; Hucho, F.
Proc. Natl. Acad. Sci. U.S.A. 88, 6117-6121, 1991
A;Title: Anionic subsites of the catalytic center of acetylcholinesterase from Torpedo a
A;Reference number: A41117; MUID:91296772
A;Accession: A41117
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-8 <KRE>
C;Keywords: carboxylic ester hydrolase


  Query Match             31.1%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 GADL 7
              ||::
Db          1 GAEM 4


RESULT  25
PN0043
phosphatidylethanol amine-binding protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 29-Oct-1997 #sequence_revision 29-Oct-1997 #text_change 23-Jan-1998
C;Accession: PN0043
R;Kato, H.
Kawasaki Igakkaishi 22, 245-259, 1996
A;Title: Analysis of proteins isolated by two dimensional electrophoresis of mouse neur
A;Reference number: PN0041
A;Accession: PN0043
A;Molecule type: protein
A;Residues: 1-8 <KAT>
A;Experimental source: neuroblastoma cell
C;Comment: The molecular mass is 23,500 and the pI is 4.8. The amino-terminus is blocked
C;Keywords: brain


  Query Match             31.1%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LSG 4
              |||
Db          5 LSG 7


RESULT  26
PT0557
```

```
T-cell receptor beta chain V-D-J region (126-1BD) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0557
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0557
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-8 <FEE>
A;Experimental source: day 18 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 SGAD 6
              || |
Db          2 SGDD 5


RESULT  27
A61364
isotocin - common carp
C;Species: Cyprinus carpio (common carp)
C;Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 20-Jun-2000
C;Accession: A61364
R;Acher, R.; Chauvet, J.; Chauvet, M.T.; Crepy, D.
Comp. Biochem. Physiol. A 14, 245-254, 1965
A;Title: Caracterisation des hormones neurohypophysaires d'un poisson osseux d'eau do
A;Reference number: A61364
A;Accession: A61364
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <ACH>
C;Superfamily: oxytocin-neurophysin
C;Keywords: amidated carboxyl end; neuropeptide; posterior pituitary
F;9/Modified site: amidated carboxyl end (Gly) #status experimental


  Query Match             31.1%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   28.6%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 YLSGADL 7
              |:|   :
Db          2 YISNCPI 8


RESULT  28
C57444
neuropeptide Grb-AST B3 - two-spotted cricket
C;Species: Gryllus bimaculatus (two-spotted cricket)
C;Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 26-Jan-1996
C;Accession: C57444
R;Lorenz, M.W.; Kellner, R.; Hoffmann, K.H.
J. Biol. Chem. 270, 21103-21108, 1995
A;Title: A family of neuropeptides that inhibit juvenile hormone biosynthesis in the
A;Reference number: A57444; MUID:95403341
A;Accession: C57444
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <LOR>


  Query Match             31.1%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          2 LSG 4
              |||
Db          5 LSG 7


RESULT  29
S43959
Ig mu chain V region (clone 13) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 20-Oct-1994 #sequence_revision 26-Jul-1996 #text_change 17-Mar-1999
C;Accession: S43959
R;Wagner, S.D.; Williams, G.T.; Larson, T.; Neuberger, M.S.; Kitamura, D.; Rajewsky, K.;
Nucleic Acids Res. 22, 1389-1393, 1994
A;Title: Antibodies generated from human immunoglobulin miniloci in transgenic mice.
A;Reference number: S43956; MUID:94248036
A;Accession: S43959
A;Molecule type: DNA
A;Residues: 1-4 <WAG>
C;Keywords: immunoglobulin


  Query Match             28.9%;  Score 13;  DB 2;  Length 4;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSG 4
              | :|
Db          1 YCAG 4


RESULT  30
I50210
gene c-rel protein - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 28-Feb-1997
C;Accession: I50210
R;Kabrun, N.; Bumstead, N.; Hayman, M.J.; Enrietto, P.J.
Mol. Cell. Biol. 10, 4788-4794, 1990
A;Title: Characterization of a novel promoter insertion in the c-rel locus.
A;Reference number: I50210; MUID:90355995
A;Accession: I50210
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-7 <KAB>
A;Cross-references: GB:M55577; NID:g555438; PID:g211661
C;Genetics:
A;Gene: c-rel


  Query Match             28.9%;  Score 13;  DB 2;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LSGA 5
              ::||
Db          1 MAGA 4


RESULT  31
PT0547
T-cell receptor beta chain V-D-J region (126-1AI) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0547
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0547
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-8 <FEE>
A;Experimental source: day 18 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             28.9%;  Score 13;  DB 2;  Length 8;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 SGAD 6
              | ||
Db          3 SDAD 6


RESULT  32
A44873
caldesmon - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 31-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 31-Oct-1997
C;Accession: A44873
R;Ikebe, M.; Hornick, T.
Arch. Biochem. Biophys. 288, 538-542, 1991
A;Title: Determination of the phosphorylation sites of smooth muscle caldesmon by pro
A;Reference number: A44873; MUID:91378498
A;Accession: A44873
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <IKE>
A;Experimental source: skeletal myosin
A;Note: sequence extracted from NCBI backbone (NCBIP:63199)
C;Superfamily: caldesmon


  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   33.3%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 GADLNL 9
              |: | :
Db          1 GSSLKI 6


RESULT  33
QDRB
delta sleep-inducing peptide - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 20-Jun-2000 #sequence_revision 20-Jun-2000 #text_change 20-Jun-2000
C;Accession: A01422
R;Monnier, M.; Dudler, L.; Gachter, R.; Maier, P.F.; Tobler, H.J.; Schoenenberger, G.
Experientia 33, 548-552, 1977
A;Title: The delta sleep inducing peptide (DSIP). Comparative properties of the origi
A;Reference number: A01422; MUID:77185324
A;Accession: A01422
A;Molecule type: protein
A;Residues: 1-9 <MON>
C;Comment: This peptide was obtained from dialysates of occipital venous sinus blood
 of recipient rabbits, it induces spindle and delta EEG activity and reduced motor ac
C;Superfamily: unassigned animal peptides


  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 SGAD 6
              :| |
Db          2 AGGD 5


RESULT  34
C41170
photosystem II protein psbM - Chlamydomonas reinhardtii (fragment)
C;Species: Chlamydomonas reinhardtii
C;Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 30-Sep-1993
C;Accession: C41170
```

```
R;de Vitry, C.; Diner, B.A.; Popot, J.L.
J. Biol. Chem. 266, 16614-16621, 1991
A;Title: Photosystem II particles from Chlamydomonas reinhardtii. Purification, molecula
A;Reference number: A41170; MUID:91358452
A;Accession: C41170
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <DE5>


  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGA 5
           ::||
Db       4 IAGA 7


RESULT  35
PH0935
T-cell receptor beta chain V-D-J region (clone 6) - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change 30-May-1997
C;Accession: PH0935
R;Gold, D.P.; Offner, H.; Sun, D.; Wiley, S.; Vandenbark, A.A.; Wilson, D.B.
J. Exp. Med. 174, 1467-1476, 1991
A;Title: Analysis of T cell receptor beta chains in Lewis rats with experimental allergi
A;Reference number: PH0891; MUID:92078857
A;Accession: PH0935
A;Molecule type: mRNA
A;Residues: 1-9 <GOL>
A;Experimental source: complete Freund's adjuvant-immunized lymph node
C;Keywords: T-cell receptor


  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGAD 6
           :||:
Db       6 TGAE 9


Search completed: December 16, 2000, 03:35:12
Job time: 5645 sec
```

```
                    GenCore version 4.5
             Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 03:07:17 ; Search time 107.12 Seconds
                                              (without alignments)
                                              2.873 Million cell updates/sec

Title:          US-09-529-121-2
Perfect score:  45
Sequence:       1 YLSGADLNL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters:      61695

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :      A_Geneseq_36:*
               1:  /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
               2:  /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
               3:  /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
               4:  /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
               5:  /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
               6:  /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
               7:  /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
               8:  /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
               9:  /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
               10:  /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
               11:  /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
               12:  /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
               13:  /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
               14:  /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
               15:  /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
               16:  /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
               17:  /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
               18:  /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
               19:  /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
               20:  /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
               21:  /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                %
Result          Query
  No.   Score   Match Length DB  ID                      Description
--------------------------------------------------------------------------------
    1      45   100.0      9  20  Y09526                  Carcinoembryonic a
    2      43    95.6      9  20  Y09527                  Carcinoembryonic a
    3      40    88.9      9  18  W39723                  Human carcina-embr
    4      40    88.9      9  19  W77134                  CEA synthetic pept
    5      40    88.9      9  19  W70045                  CEA derived HLA-A2
    6      40    88.9      9  20  Y47655                  Immunogenic peptid
    7      40    88.9      9  20  Y09525                  Carcinoembryonic a
    8      38    84.4      9  20  Y09528                  Carcinoembryonic a
    9      37    82.2      9  21  Y54173                  HLA binding peptid
   10      36    80.0      9  20  Y09529                  Carcinoembryonic a
   11      33    73.3      9  17  W00680                  Peptide comprising
   12      25    55.6      7  20  Y41847                  Rheumatoid arthrit
   13      25    55.6      9  18  W38383                  Synthetic pMEL17 p
   14      25    55.6      9  20  Y47062                  Immunogenic peptid
   15      24    53.3      9  17  W00690                  NCA analogue of re
   16      24    53.3      9  19  W70078                  B. stearothermophi
   17      23    51.1      7  20  Y41846                  Rheumatoid arthrit
   18      23    51.1      9  11  R07966                  Tryptic fragment T
   19      21    46.7      6  11  R09414                  LFA-1 alpha subuni
   20      21    46.7      6  17  W02264                  Gingivalis adhesio
   21      21    46.7      6  17  R88476                  Internal tryptic p
   22      21    46.7      9  19  W40267                  K. oxytoca R-speci
   23      21    46.7      9  20  Y47818                  Immunogenic peptid
   24      20    44.4      6  19  W56879                  Enzyme inhibitor p
   25      20    44.4      7  20  Y41962                  Rheumatoid arthrit
   26      20    44.4      7  20  Y42028                  Rheumatoid arthrit
   27      20    44.4      8  16  R73336                  Human TSH receptor
   28      20    44.4      8  19  W56983                  Enzyme inhibitor p
   29      20    44.4      9  15  R59233                  Peptide fragment (
   30      20    44.4      9  16  R70067                  Control peptide 92
   31      20    44.4      9  16  R67605                  Jojoba fatty acyl-
   32      20    44.4      9  16  R67613                  Jojoba fatty acyl-
   33      20    44.4      9  19  W70077                  Thermus thermophil
   34      20    44.4      9  19  W54515                  Synthetic polypept
   35      20    44.4      9  21  Y70832                  BH3 domain of mous
   36      19    42.2      7  13  R22435                  Amino acids coded
   37      19    42.2      8  15  R46699                  Monomeric peptide
   38      19    42.2      9  14  R30155                  MAb GAH variable r
   39      19    42.2      9  16  R87430                  Human MHC class II
   40      19    42.2      9  19  W54298                  Human cytohesin-1
   41      19    42.2      9  20  Y55448                  HLA binding plu-1
   42      19    42.2      9  20  Y55529                  HLA binding plu-1
   43      19    42.2      9  20  Y55623                  HLA binding plu-1
   44      19    42.2      9  20  Y47532                  Immunogenic peptid
   45      19    42.2      9  21  Y56575                  Tyrosinase 2 deriv
   46      18    40.0      5  19  W31459                  Transcriptional ac
   47      18    40.0      5  20  Y28182                  Factor Xa inhibiti
   48      18    40.0      6  19  W83890                  Peptide specific a
   49      18    40.0      7  15  R46809                  Phytase derived pe
   50      18    40.0      7  19  W69333                  Haemoglobin mutant
   51      18    40.0      7  20  Y17023                  Heat shock protein
   52      18    40.0      7  20  Y05043                  Tumour antigen ant
   53      18    40.0      7  21  Y85161                  Trehalose-releasin
   54      18    40.0      8  10  P91779                  Synthetic SCM-acti
   55      18    40.0      8  16  R74536                  Protease Nexin-1 P
   56      18    40.0      8  19  W53313                  CS4-CFA/I family s
   57      18    40.0      8  20  Y16852                  Heat shock protein
   58      18    40.0      8  20  W94102                  VIF-derived HIV pr
   59      18    40.0      8  20  W82429                  PVX-2 viral replic
   60      18    40.0      9  13  R22425                  Peptide derived fr
   61      18    40.0      9  14  R46546                  Jardetzky self pep
   62      18    40.0      9  15  Y38246                  HIV-derived HLA-bi
   63      18    40.0      9  16  R87433                  Human MHC class II
   64      18    40.0      9  17  Y07297                  Peptide against SH
   65      18    40.0      9  17  W49278                  Human leucocyte an
   66      18    40.0      9  19  W78574                  SH2 domain binding
   67      18    40.0      9  19  W78670                  SH2 domain binding
   68      18    40.0      9  20  Y55545                  HLA binding plu-1
   69      18    40.0      9  20  Y55587                  HLA binding plu-1
   70      18    40.0      9  20  Y41907                  Rheumatoid arthrit
   71      18    40.0      9  20  Y42029                  Rheumatoid arthrit
   72      18    40.0      9  20  Y42038                  Rheumatoid arthrit
   73      18    40.0      9  20  Y48656                  Membrane dipeptida
   74      18    40.0      9  20  Y31774                  Mycobacterial hsp6
   75      18    40.0      9  20  Y45818                  Immunogenic peptid

                               ALIGNMENTS

RESULT   1
Y09526
ID   Y09526 standard; peptide; 9 AA.
XX
AC   Y09526;
```

```
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:2.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             100.0%;  Score 45;  DB 20;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
           |||||||||
Db       1 ylsgadlnl 9

RESULT   2
Y09527
ID   Y09527 standard; peptide; 9 AA.
XX
AC   Y09527;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:3.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             95.6%;  Score 43;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
           ||||||:||
Db       1 ylsgadinl 9

RESULT   3
W39723
ID   W39723 standard; peptide; 9 AA.
XX
AC   W39723;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Human carcina-embryonic antigen (CEA) peptide (pos. 571-579).
XX
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
XX
OS   Homo sapiens.
XX
PN   WO9741440-A1.
XX
PD   06-NOV-1997.
XX
PF   28-APR-1997;   97WO-NL00229.
XX
PR   23-DEC-1996;   96EP-0203670.
PR   26-APR-1996;   96EP-0201145.
XX
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
```

```
XX
PI   Kast WM,  Melief CJM,  Offringa R,  Toes REM,  Van Der Burg SH;
XX
DR   WPI; 1997-549891/50.
XX
PT   Method of selecting T cell peptide epitope(s) - by measuring the
PT   stability of HLA class I-peptide complexes on intact B cells
XX
PS   Example 3; Page 85; 109pp; English.
XX
CC   Peptides W39430-W39734 are used in a novel method for the selection of
CC   immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC   method involves the identification of peptide sequences capable of
CC   binding to an HLA (human leukocyte antigen) class I molecule and
CC   measuring the binding of this epitope peptide to the HLA class I peptide.
CC   The stability of binding of the peptide and MHC (major histocompatablilty
CC   complex) class I molecule is measured on intact human B cells carrying
CC   the MHC molecule at their cell surfaces. The method can be used to select
CC   peptide epitopes for generating vaccines against a disease associated
CC   with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC   especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC   immune responses. Peptide W39723 is derived from the human
CC   carcino-embryonic antigen (CEA) and has the ability to bind to the human
CC   MHC Class I allele HLA-A2.1.
XX
SQ   Sequence   9 AA;


  Query Match             88.9%;  Score 40;  DB 18;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADLNL 9
              |||||:|||
Db        1 ylsganlnl 9


RESULT   4
W77134
ID   W77134 standard; peptide; 9 AA.
XX
AC   W77134;
XX
DT   16-NOV-1998  (first entry)
XX
DE   CEA synthetic peptide epitope 1.
XX
KW   Tyrosinase; tyrosinase cytotoxic lymphocyte response;
KW   cytotoxic T lymphocyte; cysteine-depleted; melanoma.
XX
OS   Synthetic.
XX
PN   WO9833810-A2.
XX
PD   06-AUG-1998.
XX
PF   29-JAN-1998;   98WO-US01592.
XX
PR   30-JAN-1997;   97US-0037781.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI   Engelhard VH,  Hunt DF,  Kittlesen D,  Slingluff CL;
XX
DR   WPI; 1998-437388/37.
XX
PT   Disease specific immunogen  - comprises disease specific cytotoxic T
PT   lymphocyte epitope used to elicit melanoma specific CTL response
XX
PS   Disclosure; Page 27; 93pp; English.
XX
CC   The peptide epitope W77119-W77138 were created for human tumour-specific
CC   cytotoxic T lymphocyte response.  These peptides are are cysteine-
```

```
CC   depleted mutants of a native disease-specific CTL epitope.  The cysteine-
CC   depleted CTL epitopes elicit a stronger or more specific CTL response
CC   than the native epitope.  The epitopes can be used in a disease-specific
CC   immunogen to protect a mammal against disease in particular melanomas.
CC   The peptides may also be used to screen a sample for the presence of
CC   an antigen with the same epitope, or with a different cross-reactive
CC   epitope.
XX
SQ   Sequence   9 AA;


  Query Match             88.9%;  Score 40;  DB 19;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADLNL 9
              |||||:|||
Db        1 ylsganlnl 9


RESULT   5
W70045
ID   W70045 standard; peptide; 9 AA.
XX
AC   W70045;
XX
DT   22-OCT-1998  (first entry)
XX
DE   CEA derived HLA-A2.1 binding peptide 2 (residues 605-613).
XX
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW   antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW   viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW   fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9833888-A1.
XX
PD   06-AUG-1998.
XX
PF   30-JAN-1998;   98WO-US01959.
XX
PR   31-JAN-1997;   97US-0036696.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Celis E,  Sette A,  Sidney J,  Southwood S,  Tsai V;
XX
DR   WPI; 1998-437445/37.
XX
PT   Production of antigen-specific cytotoxic T cells - by incubating
PT   immunogenic peptide(s) from antigen that binds class I major
PT   histocompatibility complex molecules with pre-treated antigen
PT   presenting cells
XX
PS   Example 6; Page 75; 104pp; English.
XX
CC   Sequences shown in W70044 to W70052 represent peptides derived from
CC   carcinoembryonic antigen (CEA). The peptides can bind to a human
CC   leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method
CC   of invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC   vitro. The method comprises contacting immunogenic peptides from an
CC   antigen that binds class I major histocompatibility complex (MHC)
CC   molecules with antigen presenting cells (APCs) pretreated with
CC   pretreatment growth factors, and incubating the APCs with purified CD8
CC   cells in the presence of at least 2 incubation growth factors, thereby
CC   producing antigen-specific CTLs. A method for specifically killing
CC   target cells in a human patient is also provided which comprises
CC   obtaining a fluid sample containing CTLs from a patient, contacting the
CC   cytotoxic T cells with APCs pretreated with pre-treatment growth factors,
```

```
CC   where the APCs comprise class I MHC molecules. The pretreated APCs are
CC   incubated with the cytotoxic growth factors, thereby producing activated
CC   CTLs which are contacted with a carrier to form a composition. The
CC   composition can then be administered to the patient. The activated CTLs
CC   can be used for treating cancers, immune disorders, viral infections,
CC   AIDS, hepatitis, bacterial infection, fungal infection, malaria or
CC   tuberculosis.
XX
SQ   Sequence   9 AA;

  Query Match              88.9%;  Score 40;  DB 19;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADLNL 9
            |||||:|||
Db        1 ylsganlnl 9

RESULT   6
Y47655
ID   Y47655 standard; Peptide; 9 AA.
XX
AC   Y47655;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2266.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 118; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
```

```
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;

  Query Match              88.9%;  Score 40;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGADLNL 9
            |||||:|||
Db        1 ylsganlnl 9

RESULT   7
Y09525
ID   Y09525 standard; peptide; 9 AA.
XX
AC   Y09525;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist CAP-1.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 1; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match              88.9%;  Score 40;  DB 20;  Length 9;
```

```
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
           |||||:|||
Db       1 ylsganlnl 9


RESULT   8
Y09528
ID   Y09528 standard; peptide; 9 AA.
XX
AC   Y09528;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:4.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;


  Query Match             84.4%;  Score 38;  DB 20;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
           |||||::||
Db       1 ylsganinl 9


RESULT   9
Y54173
ID   Y54173 standard; peptide; 9 AA.
XX
AC   Y54173;
XX
DT   06-APR-2000  (first entry)
XX
DE   HLA binding peptide 1233.11 derived from source CEA.605V9.
XX
KW   Allele-specific binding motif; major histocompatibility complex; MHC;
KW   HLA; HLA-A2.1; cytotoxic T cell response; antigen; microbial infection;
KW   hepatitis; AIDS; malaria; condyloma acuminatum; cancer; prostate cancer;
KW   renal carcinoma; cervical carcinoma; lymphoma; tumour.
XX
OS   Unidentified.
XX
PN   WO9965522-A1.
XX
PD   23-DEC-1999.
XX
PF   17-JUN-1999;   99WO-US13789.
XX
PR   17-JUN-1998;   98US-0098584.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Sidney J,  Southwood S;
XX
DR   WPI; 2000-106018/09.
XX
PT   Novel HLA binding immunogenic peptides used to induce T cell activation
PT   and to induce an immune response  -
XX
PS   Claim 1; Page 32; 42pp; English.
XX
CC   Peptides Y54171-Y54236 represent immunogenic peptides comprising an
CC   allele-specific binding motif for the major histocompatibility complex
CC   (MHC) molecule HLA, e.g. HLA-A2.1. The peptides have conserved residues
CC   at certain positions such as positions 2 and 9. Also, the peptides do not
CC   comprise negative binding residues at other positions, such as positions
CC   1, 3, 6 and/or 7 (peptides 9 amino acids long) and at poisons 1, 3, 4,
CC   5, 7, 8 and/or 9 (peptides 10 amino acids long). The peptides are used to
CC   induce a cytotoxic T cell response to a preselected antigen. The method
CC   comprises contacting cytotoxic T cells from a patient (optionally
CC   expressing a specific MHC class I allele) with the present peptides.
CC   The peptides are used to treat and prevent microbial infection (e.g. in
CC   viral hepatitis B and C, human papilomavirus (HPV) infection, AIDS,
CC   cytomegalovirus (CMV), malaria, and condyloma acuminatum) and cancer
CC   (e.g. prostate cancer, renal carcinoma, cervical carcinoma, lymphoma).
CC   Patients in the acute phase of infection can be treated with the
CC   peptides in conjunction with other treatments. The antigenic peptides
CC   may be used to elicit cytotoxic T lymphocytes (CTLs) ex vivo and in
CC   vivo. The resulting CTLs can be used to treat chronic infections (viral
CC   or bacterial) or tumours in patients that do not respond to conventional
CC   forms of therapy. The peptides may also be used to produce monoclonal
CC   antibodies, which are useful as potential diagnostic or therapeutic
CC   agents. The peptides may also be used as diagnostic reagents.
XX
SQ   Sequence   9 AA;


  Query Match             82.2%;  Score 37;  DB 21;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGADLNL 9
           |||||:||:
Db       1 ylsganlnv 9


RESULT  10
Y09529
ID   Y09529 standard; peptide; 9 AA.
XX
AC   Y09529;
```

```
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:5.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;


  Query Match             80.0%;  Score 36;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGADLNL 9
            ||||| |||
Db        1 ylsgaclnl 9


RESULT  11
W00680
ID   W00680 standard; peptide; 9 AA.
XX
AC   W00680;
XX
DT   01-MAY-1997  (first entry)
XX
DE   Peptide comprising residues 571-579 of Carcinoembryonic antigen.
XX
KW   Carcinoembryonic; antigen; human; cytotoxic T cell; pox virus;
KW   vector; epitope; determination; screening; tumour; treatment.
XX
OS   Homo sapiens.
XX
PN   WO9626271-A1.
```

```
XX
PD   29-AUG-1996.
XX
PF   13-FEB-1996;   96WO-US02156.
XX
PR   22-FEB-1995;   95US-0396385.
XX
PA   (THER-) THERION BIOLOGICS CORP.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Panicali D,  Schlom J,  Tsang KY;
XX
DR   WPI; 1996-402364/40.
XX
PT   Generation of human cytotoxic T-cells specific for CEA - useful in
PT   therapy, epitope mapping and drug screening
XX
PS   Claim 4; Page 57; 76pp; English.
XX
CC   Producing carcinoembryonic antigen (CEA) specific human cytotoxic T
CC   cells (CTC), comprises introducing a 1st pox virus vector, having
CC   at least 1 insertion site containing a DNA segment encoding a CEA
CC   peptide (i.e. the present peptide) to a host to stimulate CTC
CC   production, and at least 1 periodic interval after that, contacting
CC   the host with an additional antigen. The CEA specific CTC can be
CC   used to determine the CTC eliciting epitope of CEA, and to screen
CC   for compounds which enhance the ability of the antigen to create a
CC   CTC response. A host with a CEA expressing tumour can be treated by
CC   introducing the CTC to the host, and at least 1 periodic interval
CC   after that introducing a CEA peptide, i.e. the present peptide.
CC   The present peptide is positive for binding to HLA-A2, and scored
CC   561 and 806 in T2 cell binding assays, where the binding of an
CC   appropriate peptide results in the upregulation of surface HLA-A2
CC   on the T2 cells, which can be quantified via FACScan using an
CC   anti-HLA-A2 antibody (background 280 and 300).
XX
SQ   Sequence   9 AA;


  Query Match             73.3%;  Score 33;  DB 17;  Length 9;
  Best Local Similarity   87.5%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSGADLNL 9
            ||||:|||
Db        2 lsganlnl 9


RESULT  12
Y41847
ID   Y41847 standard; Peptide; 7 AA.
XX
AC   Y41847;
XX
DT   09-DEC-1999  (first entry)
XX
DE   Rheumatoid arthritis diagnostic protein isoform 17 peptide #2.
XX
KW   Human; rheumatoid arthritis; RA; diagnosis; RPI; RADF; detection;
KW   rheumatoid arthritis diagnostic feature; ERPI; synovial fluid;
KW   rheumatoid arthritis diagnostic protein isoform; screening;
KW   expression reference protein isoform; prognosis.
XX
OS   Homo sapiens.
XX
PN   WO9947925-A2.
XX
PD   23-SEP-1999.
XX
PF   15-MAR-1999;   99WO-GB00763.
XX
PR   13-MAR-1998;   98GB-0005477.
XX
```

```
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Parekh RB,  Patel TP,  Townsend RR;
XX
DR   WPI; 1999-571871/48.
XX
PT   Diagnosis of human rheumatoid arthritis by two-dimensional
PT   electrophoresis -
XX
PS   Claim 20; Page 150; 157pp; English.
XX
CC   A method has been developed for the diagnosis of human rheumatoid
CC   arthritis (RA) using two-dimensional electrophoresis to generate a
CC   two-dimensional array of features. The method can be used for screening,
CC   diagnosis and prognosis of RA in a subject or for monitoring the effect
CC   of an anti-RA drug or therapy administered to a subject. The method
CC   comprises: (a) analysing a sample of serum or plasma and optionally
CC   synovial fluid by two-dimensional electrophoresis, to generate a two-
CC   dimensional array of features; (b) identifying at least one chosen
CC   feature whose relative abundance correlates with the presence or absence
CC   of RA; and (c) comparing the abundance of each chosen feature in the
CC   sample with the abundance of that chosen feature in serum or plasma from
CC   one or more persons without RA, where the relative abundance of the
CC   chosen feature or features in the sample indicates the presence or
CC   absence of RA in the subject. The method can also be used in clinical
CC   studies for testing drugs for therapy of RA, for purification of RA-
CC   diagnostic protein isoforms (RPIs), and for production of antibodies to
CC   RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify
CC   compounds that promote or inhibit their activity, which are then used as
CC   RA drugs. Nucleic acid encoding RADFs can be used in gene therapy
CC   protocols. Y41844 to Y42100 represent RPI peptides, Y42101 to Y42103
CC   represent expression reference protein isoform peptides and Z25066 to
CC   Z25068 represent degenerate probes for RPIs, which are all used in
CC   the exemplification of the present invention.
XX
SQ   Sequence   7 AA;


  Query Match             55.6%;  Score 25;  DB 20;  Length 7;
  Best Local Similarity   83.3%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGADLN 8
           |||||:
Db       2 sgadls 7


RESULT  13
W38383
ID   W38383 standard; peptide; 9 AA.
XX
AC   W38383;
XX
DT   08-APR-1998  (first entry)
XX
DE   Synthetic pMEL17 peptide.
XX
KW   Melanoma; immunogen; cytotoxic T lymphocyte; CTL;
KW   human leukocyte antigen-A1; HLA-A1; human leukocyte antigen-A3;
KW   HLA-A3; epitope; pMEL-17; tyrosinase; vaccine; protection.
XX
OS   Synthetic.
XX
PN   WO9734613-A1.
XX
PD   25-SEP-1997.
XX
PF   17-MAR-1997;   97WO-US04958.
XX
PR   04-OCT-1996;   96US-0027627.
PR   19-MAR-1996;   96US-0013972.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
```

```
XX
PI   Cox AL,  Engelhard VH,  Hendrikson RC,  Hunt DF,  Kittlesen D;
PI   Shabanowitz J,  Skipper J,  Slingluff CL;
XX
DR   WPI; 1997-479982/44.
XX
PT   Melanoma-specific immunogens of pMel-17 and tyrosinase - useful in
PT   vaccination for producing melanoma-specific cytotoxic T lymphocytes
XX
PS   Example 9; Page 65; 106pp; English.
XX
CC   The present peptide was used in the preparation of a novel melanoma
CC   specific immunogen, comprising at least 1 melanoma specific
CC   cytotoxic T lymphocyte (CTL) epitope, where at least 1 of the
CC   epitopes is substantially homologous to a human leukocyte
CC   antigen-A1 (HLA-A1) and HLA-A3 restricted epitope of a melanoma
CC   antigen, either pMEL-17 or tyrosinase. The immunogen can be used in
CC   vaccines for protection against melanoma in mammals.
XX
SQ   Sequence   9 AA;


  Query Match             55.6%;  Score 25;  DB 18;  Length 9;
  Best Local Similarity   62.5%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGADLN 8
             ||: |||:
Db       1 ylaeadls 8


RESULT  14
Y47062
ID   Y47062 standard; Peptide; 9 AA.
XX
AC   Y47062;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #1673.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 92; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
```

```
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;


  Query Match             55.6%;  Score 25;  DB 20;  Length 9;
  Best Local Similarity   62.5%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         1 YLSGADLN 8
             ||: |||:
Db         1 ylaeadls 8


RESULT  15
W00690
ID   W00690 standard; peptide; 9 AA.
XX
AC   W00690;
XX
DT   01-MAY-1997  (first entry)
XX
DE   NCA analogue of residues 571-579 of carcinoembryonic antigen.
XX
KW   Carcinoembryonic; antigen; epitope; NCA; analogue.
XX
OS   Homo sapiens.
XX
PN   WO9626271-A1.
XX
PD   29-AUG-1996.
XX
PF   13-FEB-1996;   96WO-US02156.
XX
PR   22-FEB-1995;   95US-0396385.
XX
PA   (THER-) THERION BIOLOGICS CORP.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Panicali D,  Schlom J,  Tsang KY;
XX
DR   WPI; 1996-402364/40.
XX
PT   Generation of human cytotoxic T-cells specific for CEA - useful in
PT   therapy, epitope mapping and drug screening
XX
PS   Example 2; Page 60; 76pp; English.
XX
CC   The present peptide is negative for binding to HLA-A2, and scored
CC   252 and 225 in T2 cell binding assays, where the binding of an
CC   appropriate peptide results in the upregulation of surface HLA-A2
CC   on the T2 cells, which can be quantified via FACScan using an
CC   anti-HLA-A2 antibody (background 280 and 300).
XX
SQ   Sequence   9 AA;


  Query Match             53.3%;  Score 24;  DB 17;  Length 9;
  Best Local Similarity   55.6%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         1 YLSGADLNL 9
             |  | :|||
Db         1 yrpgenlnl 9


RESULT  16
W70078
ID   W70078 standard; peptide; 9 AA.
XX
AC   W70078;
XX
DT   28-OCT-1998  (first entry)
XX
DE   B. stearothermophilus methionyl-tRNA synthetase fragment.
XX
KW   Mycobacterium tuberculosis; mycobacterial; methionyl-tRNA synthetase;
KW   enzyme; antibiotic; tuberculosis; Bacillus stearothermophilus.
XX
OS   Bacillus stearothermophilus.
XX
PN   US5798240-A.
XX
PD   25-AUG-1998.
XX
PF   11-JAN-1996;   96US-0584226.
XX
PR   13-SEP-1994;   94US-0305766.
PR   11-JAN-1996;   96US-0584226.
XX
PA   (CUBI-) CUBIST PHARM INC.
XX
PI   Kim S,  Lee SH,  Martinis SA,  Sassanfar M,  Schimmel PR;
XX
DR   WPI; 1998-480383/41.
XX
PT   Recombinant genes encoding mycobacterial amino acyl tRNA synthetases
PT   - useful for recombinant production of the enzyme for use in
PT   screening of antibiotics against Mycobacterium tuberculosis
XX
PS   Example 1; Columns 33-34; 32pp; English.
XX
CC   Sequences shown in W70077 to W70086 represent methionyl-tRNA (Met-tRNA)
CC   synthetase fragments from different bacterial species. These are used
CC   for designing degenerate primers (V43887 to V43890) used for isolating
CC   Met-tRNa sythetase gene fragment by PCR from M. tuberculosis and M.
CC   kansasii. The invention provides an expression vector comprising a
CC   nucleic acid encoding a mycobacterial aminoacyl-tRNAs (atRNA) (especially
CC   Met-tRNA) synthetase under control of transcriptional signals that can be
CC   used to transform suitable host cells. The nucleict acids and host cells
CC   are used for the recombinant production of mycobacterial amino acyl tRNA
CC   synthetases, especially of methionyl tRNA synthetase. The enzymes are
CC   used by the organism in protein synthesis, and as such, the recombinant
CC   enzyme can be used to identify candidate drugs for use as antibiotics
CC   towards mycobacteria, especially M. tuberculosis, which is responsible
CC   for tuberculosis. Antisense constructs of the nucleic acids can also be
CC   used in antisense inhibition of the synthetase gene. The recombinant
CC   enzyme allows quick assays in screening of antibiotics. Present testing
CC   protocols involve exposing whole mycobacteria to candidate drugs and
CC   seeing their effect. This is time consuming as the bacteria are generally
CC   slow growing. Use of the enzyme also prevents researchers from having to
CC   work with pathogenic strains.
XX
SQ   Sequence   9 AA;


  Query Match             53.3%;  Score 24;  DB 19;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy          1 YLSGAD 6
              ||:| |
Db          1 yltgtd 6


RESULT   17
Y41846
ID   Y41846 standard; Peptide; 7 AA.
XX
AC   Y41846;
XX
DT   09-DEC-1999  (first entry)
XX
DE   Rheumatoid arthritis diagnostic protein isoform 17 peptide #1.
XX
KW   Human; rheumatoid arthritis; RA; diagnosis; RPI; RADF; detection;
KW   rheumatoid arthritis diagnostic feature; ERPI; synovial fluid;
KW   rheumatoid arthritis diagnostic protein isoform; screening;
KW   expression reference protein isoform; prognosis.
XX
OS   Homo sapiens.
XX
PN   WO9947925-A2.
XX
PD   23-SEP-1999.
XX
PF   15-MAR-1999;   99WO-GB00763.
XX
PR   13-MAR-1998;   98GB-0005477.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Parekh RB,  Patel TP,  Townsend RR;
XX
DR   WPI; 1999-571871/48.
XX
PT   Diagnosis of human rheumatoid arthritis by two-dimensional
PT   electrophoresis -
XX
PS   Claim 20; Page 150; 157pp; English.
XX
CC   A method has been developed for the diagnosis of human rheumatoid
CC   arthritis (RA) using two-dimensional electrophoresis to generate a
CC   two-dimensional array of features. The method can be used for screening,
CC   diagnosis and prognosis of RA in a subject or for monitoring the effect
CC   of an anti-RA drug or therapy administered to a subject. The method
CC   comprises: (a) analysing a sample of serum or plasma and optionally
CC   synovial fluid by two-dimensional electrophoresis, to generate a two-
CC   dimensional array of features; (b) identifying at least one chosen
CC   feature whose relative abundance correlates with the presence or absence
CC   of RA; and (c) comparing the abundance of each chosen feature in the
CC   sample with the abundance of that chosen feature in serum or plasma from
CC   one or more persons without RA, where the relative abundance of the
CC   chosen feature or features in the sample indicates the presence or
CC   absence of RA in the subject. The method can also be used in clinical
CC   studies for testing drugs for therapy of RA, for purification of RA-
CC   diagnostic protein isoforms (RPIs), and for production of antibodies to
CC   RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify
CC   compounds that promote or inhibit their activity, which are then used as
CC   RA drugs. Nucleic acid encoding RADFs can be used in gene therapy
CC   protocols. Y41844 to Y42100 represent RPI peptides, Y42101 to Y42103
CC   represent expression reference protein isoform peptides and Z25066 to
CC   Z25068 represent degenerate probes for RPIs, which are all used in
CC   the exemplification of the present invention.
XX
SQ   Sequence   7 AA;


  Query Match             51.1%;  Score 23;  DB 20;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 SGADLN 8
```

```
              ||||::
Db          2 sgadis 7
            -

RESULT   18
R07966
ID   R07966 standard; protein; 9 AA.
XX
AC   R07966;
XX
DT   14-JAN-1991  (first entry)
XX
DE   Tryptic fragment T32 of human N-lipocortin.
XX
KW   Human N-lipocortin; placenta; inflammation reduction; arthritis;
KW   phospholipase A2 inhibitor; tryptic fragment T32.
XX
OS   Homo sapiens.
XX
PN   US4950646-A.
XX
PD   21-AUG-1990.
XX
PF   10-JAN-1986;   86US-0929199.
XX
PR   10-JAN-1986;   86US-0929199.
PR   05-SEP-1985;   85US-0772892.
PR   14-AUG-1985;   85US-0765877.
PR   15-MAR-1985;   85US-0712376.
PR   10-JAN-1985;   85US-0690146.
XX
PA   (BIOJ ) BIOGEN NV.
XX
PI   Wallner BP,  Pepinsky RB,  Garwin JL,  Schindler DG,  Huang KS;
XX
DR   WPI; 1990-274549/36.
XX
PT   Pure fragment of human lipocortin - useful for reducing
PT   inflammation or for treating arthritis, etc.
XX
PS   Disclosure; Fig 25; 51pp; English.
XX
CC   T32 corresponds to a peak from the tryptic map of N-lipocortin,
CC   isolated from human placenta, on a Speed Vac Concentration.
CC   Amino acids 1 can also be G.
CC   Based on the similarity in the phospholipase A2 inhibitory activity
CC   of lipocortin and N-lipocortin and the similarity in the protein and DNA
CC   sequences, it was concluded that the two proteins represent a
CC   family of related proteins.  There is ca. 60% homology.
CC   The protein can be used for reducing inflammation or treating
CC   arthritic, allergic, dermatologic, opthalmic and collagen diseases
CC   and other diseases involving inflammation processes.
CC   See also Q05805-25, Q06581, R07926-37 and R07956-66.
XX
SQ   Sequence   9 AA;


  Query Match             51.1%;  Score 23;  DB 11;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGAD 6
              || | |
Db          4 ylxggd 9


RESULT   19
R09414
ID   R09414 standard; Peptide; 6 AA.
XX
AC   R09414;
XX
DT   23-AUG-1990  (first entry)
```

```
XX
DE   LFA-1 alpha subunit polypeptide (k).
XX
KW   Lymphocyte function associated antigen; inflammation; metastasis.
XX
PN   EP362526-A.
XX
PD   11-APR-1990.
XX
PF   17-AUG-1989;   89EP-0115160.
XX
PR   23-AUG-1988;   88US-0235227.
PR   09-MAR-1989;   89US-0321017.
XX
PA   (DANA-) DANA FARBER CANCER.
XX
PI   Springer TA,  Larson R;
XX
DR   WPI; 1990-108985/15.
XX
PT   Pure alpha subunit of lymphocyte function associated antigen -
PT   and encoding DNA sequences, useful eg for suppressing
PT   inflammation or metastasis
XX
PS   Claim 6; Page 19; 27pp; English.
XX
CC   The alpha-subunit (a-SU), contg. al least one of the polypeptides
CC   given in R09404-417, can bind to ICAM-1 (or other natural
CC   ligands) on the surface of cells, and can associate with the beta-SU
CC   to form a heterodimer (also able to bind to ICAM-I).  a-SU, and its
CC   derivs., are useful in suppressing inflammation, metastasis and
CC   growth of a-SU expressing tumour cells and is used in the treatment
CC   of viral infections.
CC   The pref. dose is 1 pg - 10 mg/kg.
XX
SQ   Sequence   6 AA;

  Query Match             46.7%;  Score 21;  DB 11;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           ||||
Db       2 ylsg 5


RESULT  20
W02264
ID   W02264 standard; peptide; 6 AA.
XX
AC   W02264;
XX
DT   22-OCT-1996  (first entry)
XX
DE   Gingivalis adhesion inhibitor comprising fimbrillin residues 240-245.
XX
KW   Fimbrillin gene; inhibition; adhesion; saliva; coated surface;
KW   prevention; periodontitis; teeth; gums; dentifrices; mouthwash;
KW   vaccine.
XX
OS   Porphyromonas gingivalis.
XX
PN   US5536497-A.
XX
PD   16-JUL-1996.
XX
PF   21-DEC-1992;   92US-0994277.
XX
PR   21-DEC-1992;   92US-0994277.
XX
PA   (UYNY ) UNIV NEW YORK STATE RES FOUND.
XX
```

```
PI   Bedi GS,  Evans RT,  Genco RJ,  Sojar HT;
XX
DR   WPI; 1996-341445/34.
XX
PT   Peptide inhibitor of Porphyromonas gingivalis adhesion to saliva
PT   coated surface - useful for preventing periodontitis by application
PT   to the teeth and gums esp. in dentifrices, mouth:washes or topical
PT   formulations
XX
PS   Claim 1; Columns 17-18; 23pp; English.
XX
CC   The present peptide was prepd. by chemical synthesis, on the basis
CC   of an amino acid sequence deduced from the DNA sequence of the
CC   cloned P. gingivalis fimbrillin gene described in J. Bacteriol,
CC   170, 1658, 1988. The peptide inhibits the adhesion of P. gingivalis
CC   to saliva coated surfaces, and is therefore useful for preventing
CC   periodontitis by application to teeth and gums, esp. in
CC   dentifrices, mouthwashes or topical formulations, or by admin. as a
CC   vaccine. In an assay to determine the effect of the peptide on
CC   P. gingivalis binding to saliva coated hydroxapatite beads, the
CC   peptide resulted in a percentage binding inhibition of
CC   approx. 30 %.
XX
SQ   Sequence   6 AA;

  Query Match             46.7%;  Score 21;  DB 17;  Length 6;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGADL 7
             :||||
Db       1 ngadl 5


RESULT  21
R88476
ID   R88476 standard; peptide; 6 AA.
XX
AC   R88476;
XX
DT   30-AUG-1996  (first entry)
XX
DE   Internal tryptic peptide from Tre6P synthase (peak 29) #2.
XX
KW   Tryptic peptide; trehalose-6-phosphate synthase; M. smegmatis; probe;
KW   trehalose; transgenic plant; heparin-activated; preservation; food;
KW   antigenic determinant; yeast; TSP1; fruit; berry; puree; jelly; jam.
XX
OS   Mycobacterium smegmatis.
XX
PN   WO9600789-A1.
XX
PD   11-JAN-1996.
XX
PF   29-JUN-1995;   95WO-FI00377.
XX
PR   29-JUN-1994;   94FI-0003133.
XX
PA   (ALKO-) ALKO GROUP LTD.
XX
PI   Holmstrom K,  Londesborough J,  Mandal A,  Mantyla E;
PI   Palva ET,  Tunnela O, Welin B;
XX
DR   WPI; 1996-077499/08.
XX
PT   New transgenic plants with increase trehalose contents - prepd. by
PT   transforming plants with a trehalose-6-phosphate synthase gene fused
PT   to a non-constitutive promoter
XX
PS   Example 6; Page 36; 55pp; English.
XX
CC   The sequences given in R88473-80 are internal tryptic peptides
```

```
CC   derived from trehalose-6-phosphate (Tre6P) synthase from M.
CC   smegmatis.  Tre6P is the key enzyme in the synthesis of trehalose
CC   via Tre6P.  The aim of the invention is to produce a transgenic
CC   plant with increase trehalose content.  Tre6P in M. smegmatis is
CC   heparin-activated and was isolated  and purified.    These peptides
CC   were derived from a protein which was purified with a mol. wt. of 55
CC   kD which  shared antigenic determinants with the yeast Tre6P
CC   synthase protein.  Using these peptides probes may be designed for
CC   the isolation of the Tre6P gene (TSP1) for the production of the
CC   transgenic plants.  The trehalose may be isolated from the transgenic
CC   plants and used in bulk preparation including the preservation of the
CC   flavour and structure of food stuffs during drying.  Fruits and berries
CC   form the transformed plants may be processed into purees, jellies and
CC   jams which have a fresher and richer flavour due to the increased
CC   trehalose content.
XX
SQ   Sequence   6 AA;

  Query Match             46.7%;  Score 21;  DB 17;  Length 6;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGA 5
           || ||
Db       1 ylega 5


RESULT  22
W40267
ID   W40267 standard; Protein; 9 AA.
XX
AC   W40267;
XX
DT   16-JUN-1998  (first entry)
XX
DE   K. oxytoca R-specific amidohydrolase peptide T5.
XX
KW   R-specific amidohydrolase; hydrolysis; racemate; nitrogen source;
KW   3,3,3-trifluoro-2-hydroxy-2-methyl propionamide; isoform.
XX
OS   Klebsiella oxytoca.
XX
PN   WO9801568-A2.
XX
PD   15-JAN-1998.
XX
PF   10-JUL-1997;   97WO-EP03670.
XX
PR   03-MAR-1997;   97CH-0000500.
PR   10-JUL-1996;   96CH-0001723.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Brieden W,  Naughton A,  Robins K,  Shaw N,  Tinschert A;
PI   Zimmermann T;
XX
DR   WPI; 1998-101063/09.
XX
PT   Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation
PT   - by stereoselective hydrolysis of corresponding racemic amide using
PT   microorganism or derived enzyme, used as drug intermediate
XX
PS   Example 10.2; Page 29; 68pp; German.
XX
CC   W40264-W40275 are peptide fragments of an R-specific amidohydrolase
CC   isolated from Klebsiella oxytoca strain PRS1. This enzyme allows the
CC   microorganism to utilise 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
CC   as its sole nitrogen source. This amidohydrolase is used in a process for
CC   preparing (R)-isoforms of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
CC   which is cheaper than prior art optical resolution of the racemate using
CC   dimethoxy strychnine or (S)-(-)- alpha -methylbenzylamine.
XX
```

```
SQ   Sequence   9 AA;

  Query Match             46.7%;  Score 21;  DB 19;  Length 9;
  Best Local Similarity   62.5%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 YLSGADLN 8
           |  || ||
Db       1 ytvgamln 8


RESULT  23
Y47818
ID   Y47818 standard; Peptide; 9 AA.
XX
AC   Y47818;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2429.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 124; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;
```

```
  Query Match             46.7%;  Score 21;  DB 20;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         1 YLSGADL 7
             |||  |:
Db         1 ylsegdm 7


RESULT  24
W56879
ID   W56879 standard; peptide; 6 AA.
XX
AC   W56879;
XX
DT   28-JUL-1998  (first entry)
XX
DE   Enzyme inhibitor peptide SEQ ID NO:80.
XX
KW   Enzyme inhibitor; t-PA; u-PA; chymotrypsin; serine protease; active;
KW   latent; substrate subtraction phage display peptide library;
KW   identification; kinase; phosphatase; serpin.
XX
OS   Homo sapiens.
XX
PN   WO9747314-A1.
XX
PD   18-DEC-1997.
XX
PF   10-JUN-1997;   97WO-US09760.
XX
PR   10-JUN-1996;   96US-0019495.
XX
PA   (SCRI ) SCRIPPS RES INST.
XX
PI   Ke S,  Madison EL;
XX
DR   WPI; 1998-062746/06.
XX
PT   Substrate subtraction phage display peptide libraries - used to
PT   distinguish between active and latent forms of enzyme, e.g. serine
PT   protease
XX
PS   Claim 25; Page 67; 138pp; English.
XX
CC   The present sequence represents an enzyme inhibitor peptide used in
CC   the method of the invention to distinguish between t-PA and u-PA. The
CC   present invention describes a substrate subtraction library for the
CC   identification of peptide substrates selective between a first enzyme
CC   (E1) and a second enzyme (E2), comprising a collection different
CC   peptides, substantially lacking peptides that are effective substrates
CC   for E1. Also described are: (1) a method (M1) for identifying peptide
CC   substrates selective between a first enzyme (E1) and a second enzyme
CC   (E2); (2) a compound comprising the amino acid sequence of a peptide
CC   identified by M1; (3) a polypeptide for use as an enzyme inhibitor
CC   comprising one of 237 amino acid sequences (see W56801 to W56947, and
CC   W56949 to W57038); (4) a recombinant DNA vector comprising DNA (I)
CC   encoding a protease inhibitor including the sequence identified by the
CC   M1; (5) a prokaryotic or eukaryotic cell containing the vector of (4);
CC   (6) an antibody (Ab) immunoreactive with at least one of the peptides
CC   identified by M1; and (7) a diagnostic assay for distinguishing between
CC   active and latent forms of protease inhibitors, that uses (Ab). The
CC   library and method are used for distinguishing between active and latent
CC   forms of enzyme inhibitors, e.g. proteases, kinases and phosphatases.
CC   (Ab) are used for affinity purification of recombinant peptides and in
CC   the identification of naturally occurring protease inhibitors. Enzyme-
CC   inhibiting peptides identified can be used to treat a serpin deficiency
CC   or a disorder of serine proteases.
XX
SQ   Sequence   6 AA;
```

```
  Query Match             44.4%;  Score 20;  DB 19;  Length 6;
  Best Local Similarity   60.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         4 GADLN 8
               | |:|
Db         2 gvdmn 6


RESULT  25
Y41962
ID   Y41962 standard; Peptide; 7 AA.
XX
AC   Y41962;
XX
DT   09-DEC-1999  (first entry)
XX
DE   Rheumatoid arthritis diagnostic protein isoform peptide #113.
XX
KW   Human; rheumatoid arthritis; RA; diagnosis; RPI; RADF; detection;
KW   rheumatoid arthritis diagnostic feature; ERPI; synovial fluid;
KW   rheumatoid arthritis diagnostic protein isoform; screening;
KW   expression reference protein isoform; prognosis.
XX
OS   Homo sapiens.
XX
FH   Key                 Location/Qualifiers
FT   Misc-difference     6
FT                       /label= Ile, Leu
XX
PN   WO9947925-A2.
XX
PD   23-SEP-1999.
XX
PF   15-MAR-1999;   99WO-GB00763.
XX
PR   13-MAR-1998;   98GB-0005477.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Parekh RB,  Patel TP,  Townsend RR;
XX
DR   WPI; 1999-571871/48.
XX
PT   Diagnosis of human rheumatoid arthritis by two-dimensional
PT   electrophoresis -
XX
PS   Disclosure; Page 20; 157pp; English.
XX
CC   A method has been developed for the diagnosis of human rheumatoid
CC   arthritis (RA) using two-dimensional electrophoresis to generate a
CC   two-dimensional array of features. The method can be used for screening,
CC   diagnosis and prognosis of RA in a subject or for monitoring the effect
CC   of an anti-RA drug or therapy administered to a subject. The method
CC   comprises: (a) analysing a sample of serum or plasma and optionally
CC   synovial fluid by two-dimensional electrophoresis, to generate a two-
CC   dimensional array of features; (b) identifying at least one chosen
CC   feature whose relative abundance correlates with the presence or absence
CC   of RA; and (c) comparing the abundance of each chosen feature in the
CC   sample with the abundance of that chosen feature in serum or plasma from
CC   one or more persons without RA, where the relative abundance of the
CC   chosen feature or features in the sample indicates the presence or
CC   absence of RA in the subject. The method can also be used in clinical
CC   studies for testing drugs for therapy of RA, for purification of RA-
CC   diagnostic protein isoforms (RPIs), and for production of antibodies to
CC   RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify
CC   compounds that promote or inhibit their activity, which are then used as
CC   RA drugs. Nucleic acid encoding RADFs can be used in gene therapy
CC   protocols. Y41844 to Y42100 represent RPI peptides, Y42101 to Y42103
CC   represent expression reference protein isoform peptides and Z25066 to
CC   Z25068 represent degenerate probes for RPIs, which are all used in
CC   the exemplification of the present invention.
```

```
XX
SQ   Sequence   7 AA;


  Query Match             44.4%;  Score 20;  DB 20;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 SGAD 6
            ||||
Db        2 sgad 5


RESULT  26
Y42028
ID   Y42028 standard; Peptide; 7 AA.
XX
AC   Y42028;
XX
DT   09-DEC-1999  (first entry)
XX
DE   Rheumatoid arthritis diagnostic protein isoform peptide #179.
XX
KW   Human; rheumatoid arthritis; RA; diagnosis; RPI; RADF; detection;
KW   rheumatoid arthritis diagnostic feature; ERPI; synovial fluid;
KW   rheumatoid arthritis diagnostic protein isoform; screening;
KW   expression reference protein isoform; prognosis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 6
FT                   /label= Ile, Leu
XX
PN   WO9947925-A2.
XX
PD   23-SEP-1999.
XX
PF   15-MAR-1999;   99WO-GB00763.
XX
PR   13-MAR-1998;   98GB-0005477.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Parekh RB,  Patel TP,  Townsend RR;
XX
DR   WPI; 1999-571871/48.
XX
PT   Diagnosis of human rheumatoid arthritis by two-dimensional
PT   electrophoresis -
XX
PS   Disclosure; Page 21; 157pp; English.
XX
CC   A method has been developed for the diagnosis of human rheumatoid
CC   arthritis (RA) using two-dimensional electrophoresis to generate a
CC   two-dimensional array of features. The method can be used for screening,
CC   diagnosis and prognosis of RA in a subject or for monitoring the effect
CC   of an anti-RA drug or therapy administered to a subject. The method
CC   comprises: (a) analysing a sample of serum or plasma and optionally
CC   synovial fluid by two-dimensional electrophoresis, to generate a two-
CC   dimensional array of features; (b) identifying at least one chosen
CC   feature whose relative abundance correlates with the presence or absence
CC   of RA; and (c) comparing the abundance of each chosen feature in the
CC   sample with the abundance of that chosen feature in serum or plasma from
CC   one or more persons without RA, where the relative abundance of the
CC   chosen feature or features in the sample indicates the presence or
CC   absence of RA in the subject. The method can also be used in clinical
CC   studies for testing drugs for therapy of RA, for purification of RA-
CC   diagnostic protein isoforms (RPIs), and for production of antibodies to
CC   RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify
CC   compounds that promote or inhibit their activity, which are then used as
CC   RA drugs. Nucleic acid encoding RADFs can be used in gene therapy
CC   protocols. Y41844 to Y42100 represent RPI peptides, Y42101 to Y42103
```

```
CC   represent expression reference protein isoform peptides and Z25066 to
CC   Z25068 represent degenerate probes for RPIs, which are all used in
CC   the exemplification of the present invention.
XX
SQ   Sequence   7 AA;


  Query Match             44.4%;  Score 20;  DB 20;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 SGAD 6
            ||||
Db        2 sgad 5


RESULT  27
R73336
ID   R73336 standard; Peptide; 8 AA.
XX
AC   R73336;
XX
DT   12-DEC-1995  (first entry)
XX
DE   Human TSH receptor (residues 271-278).
XX
KW   thyroid stimulating hormone receptor; TSH; human; Homo sapiens;
KW   antibody; affinity; detection.
XX
OS   Synthetic.
XX
PN   JP07089991-A.
XX
PD   04-APR-1995.
XX
PF   28-SEP-1993;   93JP-0240853.
XX
PR   28-SEP-1993;   93JP-0240853.
XX
PA   (MITP ) MITSUBISHI PETROCHEMICAL CO LTD.
XX
DR   WPI; 1995-167251/22.
XX
PT   Novel polypeptide(s) having affinity for the human TSH receptor
PT   antibody - used in detection of the TSH antibody.
XX
PS   Example 1; Page 24; 54pp; Japanese.
XX
CC   Peptides with affinity to human TSH (thyroid stimulating hormone)
CC   receptor antibody are used for detection of the antibody. (See also
CC   R73201-592).
XX
SQ   Sequence   8 AA;


  Query Match             44.4%;  Score 20;  DB 16;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGADLN 8
              :|: |||:
Db        1 hltradls 8


RESULT  28
W56983
ID   W56983 standard; peptide; 8 AA.
XX
AC   W56983;
XX
DT   28-JUL-1998  (first entry)
XX
DE   Enzyme inhibitor peptide SEQ ID NO:184.
```

```
XX
KW   Enzyme inhibitor; t-PA; u-PA; chymotrypsin; serine protease; active;
KW   latent; substrate subtraction phage display peptide library;
KW   identification; kinase; phosphatase; serpin.
XX
OS   Homo sapiens.
XX
PN   WO9747314-A1.
XX
PD   18-DEC-1997.
XX
PF   10-JUN-1997;   97WO-US09760.
XX
PR   10-JUN-1996;   96US-0019495.
XX
PA   (SCRI ) SCRIPPS RES INST.
XX
PI   Ke S,  Madison EL;
XX
DR   WPI; 1998-062746/06.
XX
PT   Substrate subtraction phage display peptide libraries - used to
PT   distinguish between active and latent forms of enzyme, e.g. serine
PT   protease
XX
PS   Claim 25; Page 102; 138pp; English.
XX
CC   The present sequence represents an enzyme inhibitor peptide used in
CC   the method of the invention to distinguish between t-PA and u-PA. The
CC   present invention describes a substrate subtraction library for the
CC   identification of peptide substrates selective between a first enzyme
CC   (E1) and a second enzyme (E2), comprising a collection different
CC   peptides, substantially lacking peptides that are effective substrates
CC   for E1. Also described are: (1) a method (M1) for identifying peptide
CC   substrates selective between a first enzyme (E1) and a second enzyme
CC   (E2); (2) a compound comprising the amino acid sequence of a peptide
CC   identified by M1; (3) a polypeptide for use as an enzyme inhibitor
CC   comprising one of 237 amino acid sequences (see W56801 to W56947, and
CC   W56949 to W57038); (4) a recombinant DNA vector comprising DNA (I)
CC   encoding a protease inhibitor including the sequence identified by the
CC   M1; (5) a prokaryotic or eukaryotic cell containing the vector of (4);
CC   (6) an antibody (Ab) immunoreactive with at least one of the peptides
CC   identified by M1; and (7) a diagnostic assay for distinguishing between
CC   active and latent forms of protease inhibitors, that uses (Ab). The
CC   library and method are used for distinguishing between active and latent
CC   forms of enzyme inhibitors, e.g. proteases, kinases and phosphatases.
CC   (Ab) are used for affinity purification of recombinant peptides and in
CC   the identification of naturally occurring protease inhibitors. Enzyme-
CC   inhibiting peptides identified can be used to treat a serpin deficiency
CC   or a disorder of serine proteases.
XX
SQ   Sequence   8 AA;

  Query Match             44.4%;  Score 20;  DB 19;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 GADLN 8
            | |:|
Db        4 gvdmn 8

RESULT  29
R59233
ID   R59233 standard; peptide; 9 AA.
XX
AC   R59233;
XX
DT   04-MAY-1995  (first entry)
XX
DE   Peptide fragment (1.0164) of HIV binds HLA-A2.1.
XX
```

```
KW   antigen; epitope; immunogenic target protein; PSA; HBVc; HBVs; EBV;
KW   HIV1; core antigen; surface antigen; pharmaceutical composition;
KW   in vivo; ex vivo; therapeutic; diagnostic; MHC class I molecule;
KW   major histocompatability complex; HLA-A2.1; 9mer; 10mer; anchor;
KW   human leukocyte antigen.
XX
OS   Human immunodeficiency virus.
XX
PN   WO9420127-A.
XX
PD   15-SEP-1994.
XX
PF   04-MAR-1994;   94WO-US02353.
XX
PR   05-MAR-1993;   93US-0027146.
PR   04-JUN-1993;   93US-0073205.
PR   29-NOV-1993;   93US-0159184.
XX
PA   (CYTE-) CYTEL CORP.
XX
PI   Grey HM,  Kast WM,  Sette A,  Sidney J;
XX
DR   WPI; 1994-302678/37.
XX
PT   Immunogenic peptide(s) having an HLA-A2.1 binding motif - used
PT   for treatment or prophylaxis of cancer, virus infection or
PT   autoimmune diseases.
XX
PS   Example 5; Page 105; 138pp; English.
XX
CC   R59104-264 are immunogenic 9mer peptides that contain a HLA-A2.1 binding
CC   motif. These peptides bind HLA-A2.1 and have a binding affinity of at
CC   least 1% as compared to a reference peptide (R71293). R59233 has an IC50
CC   of 0 and the sequence occurs at position 614 in the HIV POL protein.
CC   The peptides of the invention can induce cytotoxic T lymphocytes which
CC   can react with target cells. They can be used for the treatment or
CC   prophlyaxis of cancer, eg. prostate cancer or lymphoma, etc.
XX
SQ   Sequence   9 AA;

  Query Match             44.4%;  Score 20;  DB 15;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 DLNL 9
            ||||
Db        1 dlnl 4

RESULT  30
R70067
ID   R70067 standard; peptide; 9 AA.
XX
AC   R70067;
XX
DT   06-OCT-1995  (first entry)
XX
DE   Control peptide 927.30 for HBV epitope stimulation of T-cells.
XX
KW   Hepatitis B virus polymerase; cytotoxic T cell response; prophylactic;
KW   vaccine; chronic; acute HBV infection; carrier.
XX
OS   Hepatitis B virus.
XX
PN   WO9503777-A.
XX
PD   09-FEB-1995.
XX
PF   01-AUG-1994;   94WO-US08685.
XX
PR   02-AUG-1993;   93US-0100870.
XX
```

```
PA   (SCRI ) SCRIPPS RES INST.
XX
PI   Chisari FV;
XX
DR   WPI; 1995-082004/11.
XX
PT   New peptides inducing cytotoxic T lymphocytes to hepatitis B
PT   virus  - are regions of HB polymerase protein, for treating acute
PT   and chronic infections
XX
PS   Example 1; Page 35; 85pp; English.
XX
CC   Peptides R70066-7 are control peptides in an assay for stimulation of
CC   cytotoxic T cells (CTLs) by epitopes from the hepatitis B virus (HBV)
CC   polymerase (HBpol) (see R70044-59).  The HBpol peptides can be used,
CC   prophylactically as vaccines, together with, or conjugated to, HBV helper
CC   epitopes (R70060-4).  The peptides can be used, particularly ex vivo, to
CC   stimulate CTL cells,which cells can be reintroduced into patients who
CC   have chronic or acute HBV infections or are carriers, especially in
CC   treatments to prevent conversion from acute to chronic infections.
XX
SQ   Sequence   9 AA;


  Query Match             44.4%;  Score 20;  DB 16;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 DLNL 9
           ||||
Db       1 dlnl 4


RESULT  31
R67605
ID   R67605 standard; Protein; 9 AA.
XX
AC   R67605;
XX
DT   14-AUG-1995  (first entry)
XX
DE   Jojoba fatty acyl-CoA reductase 56kd protein fragment.
XX
KW   Jojoba; fatty acyl CoA reductase; fatty acyl; fatty alcohol.
XX
OS   Simmondsia chinensis.
XX
PN   US5370996-A.
XX
PD   06-DEC-1994.
XX
PF   22-FEB-1991;   91US-0659975.
XX
PR   22-FEB-1991;   91US-0659975.
PR   27-SEP-1991;   91US-0767251.
PR   20-NOV-1991;   91US-0796256.
PR   31-JUL-1992;   92US-0920430.
XX
PA   (CALJ ) CALGENE INC.
XX
PI   Lassner MW,  Metz JG,  Pollard MR;
XX
DR   WPI; 1995-021884/03.
XX
PT   New recombinant constructs for transforming plants of E. coli  -
PT   contg. nucleic acid encoding jojoba embryo long chain fatty
PT   acyl-CoA reductase.
XX
PS   Example 4; Column 26; 30pp; English.
XX
CC   The sequence encoding the jojoba embryo long chain fatty acyl-CoA
CC   reductase may be used in recombinant constructs which in turn can be
CC   used to transform E. coli.  Such constructs are useful for the
```

```
CC   expression of the jojoba embryo long chain fatty acyl CoA reductase
CC   in host cells.  The enzyme catalyses the formation of a fatty
CC   alcohol from a fatty acyl substrate.  Analysis of the enzyme by SDS
CC   PAGE gave two prominent bands having apparent molecular masses of
CC   approximately 52 and 54 kD.  As the apparent size of the reductase
CC   enzyme in the native state is approximately 49 kD as determined by
CC   size exclusion chromatography, these bands probably represented two
CC   related forms of the enzyme instead of two different subunits of the
CC   enzyme. This sequence is a peptide fragment from the 56 kD protein.
XX
SQ   Sequence   9 AA;

  Query Match             44.4%;  Score 20;  DB 16;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GADLNL 9
           | |:|:
Db       2 gldinv 7


RESULT  32
R67613
ID   R67613 standard; Protein; 9 AA.
XX
AC   R67613;
XX
DT   14-AUG-1995  (first entry)
XX
DE   Jojoba fatty acyl-CoA reductase 54kd protein fragment.
XX
KW   Jojoba; fatty acyl CoA reductase; fatty acyl; fatty alcohol.
XX
OS   Simmondsia chinensis.
XX
PN   US5370996-A.
XX
PD   06-DEC-1994.
XX
PF   22-FEB-1991;   91US-0659975.
XX
PR   22-FEB-1991;   91US-0659975.
PR   27-SEP-1991;   91US-0767251.
PR   20-NOV-1991;   91US-0796256.
PR   31-JUL-1992;   92US-0920430.
XX
PA   (CALJ ) CALGENE INC.
XX
PI   Lassner MW,  Metz JG,  Pollard MR;
XX
DR   WPI; 1995-021884/03.
XX
PT   New recombinant constructs for transforming plants of E. coli  -
PT   contg. nucleic acid encoding jojoba embryo long chain fatty
PT   acyl-CoA reductase.
XX
PS   Example 4; Column 26; 30pp; English.
XX
CC   The sequence encoding the jojoba embryo long chain fatty acyl-CoA
CC   reductase may be used in recombinant constructs which in turn can be
CC   used to transform E. coli.  Such constructs are useful for the
CC   expression of the jojoba embryo long chain fatty acyl CoA reductase
CC   in host cells.  The enzyme catalyses the formation of a fatty
CC   alcohol from a fatty acyl substrate.  Analysis of the enzyme by SDS
CC   PAGE gave two prominent bands having apparent molecular masses of
CC   approximately 52 and 54 kD.  As the apparent size of the reductase
CC   enzyme in the native state is approximately 49 kD as determined by
CC   size exclusion chromatography, these bands probably represented two
CC   related forms of the enzyme instead of two different subunits of the
CC   enzyme. This sequence is a peptide fragment from the 54 kD protein.
XX
SQ   Sequence   9 AA;
```

```
  Query Match             44.4%;  Score 20;  DB 16;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 GADLNL 9
            | |:|:
Db        2 gldinv 7

RESULT  33
W70077
ID   W70077 standard; peptide; 9 AA.
XX
AC   W70077;
XX
DT   28-OCT-1998  (first entry)
XX
DE   Thermus thermophilus methionyl-tRNA synthetase fragment.
XX
KW   Mycobacterium tuberculosis; mycobacterial; methionyl-tRNA synthetase;
KW   enzyme; antibiotic; tuberculosis; Thermus thermophilus.
XX
OS   Thermus thermophilus.
XX
PN   US5798240-A.
XX
PD   25-AUG-1998.
XX
PF   11-JAN-1996;   96US-0584226.
XX
PR   13-SEP-1994;   94US-0305766.
PR   11-JAN-1996;   96US-0584226.
XX
PA   (CUBI-) CUBIST PHARM INC.
XX
PI   Kim S,  Lee SH,  Martinis SA,  Sassanfar M,  Schimmel PR;
XX
DR   WPI; 1998-480383/41.
XX
PT   Recombinant genes encoding mycobacterial amino acyl tRNA synthetases
PT   - useful for recombinant production of the enzyme for use in
PT   screening of antibiotics against Mycobacterium tuberculosis
XX
PS   Example 1; Columns 33-34; 32pp; English.
XX
CC   Sequences shown in W70077 to W70086 represent methionyl-tRNA (Met-tRNA)
CC   synthetase fragments from different bacterial species. These are used
CC   for designing degenerate primers (V43887 to V43890) used for isolating
CC   Met-tRNa sythetase gene fragment by PCR from M. tuberculosis and M.
CC   kansasii. The invention provides an expression vector comprising a
CC   nucleic acid encoding a mycobacterial aminoacyl-tRNAs (atRNA) (especially
CC   Met-tRNA) synthetase under control of transcriptional signals that can be
CC   used to transform suitable host cells. The nucleict acids and host cells
CC   are used for the recombinant production of mycobacterial amino acyl tRNA
CC   synthetases, especially of methionyl tRNA synthetase. The enzymes are
CC   used by the organism in protein synthesis, and as such, the recombinant
CC   enzyme can be used to identify candidate drugs for use as antibiotics
CC   towards mycobacteria, especially M. tuberculosis, which is responsible
CC   for tuberculosis. Antisense constructs of the nucleic acids can also be
CC   used in antisense inhibition of the synthetase gene. The recombinant
CC   enzyme allows quick assays in screening of antibiotics. Present testing
CC   protocols involve exposing whole mycobacteria to candidate drugs and
CC   seeing their effect. This is time consuming as the bacteria are generally
CC   slow growing. Use of the enzyme also prevents researchers from having to
CC   work with pathogenic strains.
XX
SQ   Sequence   9 AA;

  Query Match             44.4%;  Score 20;  DB 19;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGAD 6
            :|:| |
Db        1 fltgtd 6

RESULT  34
W54515
ID   W54515 standard; peptide; 9 AA.
XX
AC   W54515;
XX
DT   01-SEP-1998  (first entry)
XX
DE   Synthetic polypeptide HBV Pol 42-50.
XX
KW   Carbohydrate-specific; cytolytic T cell; therapeutic; tumour;
KW   major histocompatibility complex; vaccine; protective immune response;
KW   pathogenic bacteria; virus; CTL response.
XX
OS   Synthetic.
XX
PN   WO9815286-A1.
XX
PD   16-APR-1998.
XX
PF   08-OCT-1997;   97WO-US18146.
XX
PR   08-OCT-1996;   96US-0028260.
XX
PA   (LJOL-) LA JOLLA INST ALLERGY & IMMUNOLOGY.
XX
PI   Grey H;
XX
DR   WPI; 1998-240595/21.
XX
PT   Composition for stimulating carbohydrate-specific cytotoxic T
PT   lymphocytes - comprises synthetic peptide with attached carbohydrate
PT   hapten, designed to bind to class I molecule, used in vaccines
PT   against cancer or infectious disease
XX
PS   Disclosure; Page 18; 64pp; English.
XX
CC   The Peptides W54491-W54531 and W54533-W54534 are examples of polypeptides
CC   tested and used in the production of a synthetic polypeptide, for
CC   stimulating a carbohydrate-specific cytolytic T cell (CTL) response.
CC   The polypeptide should be at least 8 amino acids, 2 of which are anchor
CC   residues.  The polypeptide binds to the binding groove of a major
CC   histocompatibility complex class I molecule and a carbohydrate molecule
CC   which is linked to an internal amino acid of the polypeptide extends
CC   beyond the groove.  This stimulates T cells which are able to lyse
CC   specifically cells that express carbohydrate residues on their surface.
CC   The polypeptides are used as vaccines to generate therapeutic or
CC   protective immune responses, particularly against tumours but also
CC   against pathogenic bacteria and viruses (e.g. mycobacteria that cause
CC   leprosy and tuberculosis).
XX
SQ   Sequence   9 AA;

  Query Match             44.4%;  Score 20;  DB 19;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 DLNL 9
            ||||
Db        1 dlnl 4

RESULT  35
Y70832
ID   Y70832 standard; peptide; 9 AA.
```

```
XX
AC   Y70832;
XX
DT   31-JUL-2000  (first entry)
XX
DE   BH3 domain of mouse BAK protein.
XX
KW   Mouse; truncated BAX protein; tBAX; BAX alpha; BCL-2 family; head trauma;
KW   neuron; anti-apoptotic; cerebroprotective; neuroprotective; neuroactive;
KW   apoptosis; treatment; neurodegenerative disease; peripheral nerve injury;
KW   spinal cord injury; stroke; pro-apoptotic; PCD; programmed cell death;
KW   BH3 domain; BCL-2 homology domain; BAK protein.
XX
OS   Mus musculus.
XX
PN   WO200023083-A1.
XX
PD   27-APR-2000.
XX
PF   22-OCT-1999;   99WO-US24747.
XX
PR   22-OCT-1998;   98US-0177315.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Johnson EM,  Easton R;
XX
DR   WPI; 2000-339513/29.
XX
PT   Truncated BAX polypeptides useful for preventing apoptosis of neurons
PT   for the treatment of nervous system disorders -
XX
PS   Disclosure; Fig 3; 43pp; English.
XX
CC   The patent discloses specific truncated BAX (tBAX) proteins
CC   which inhibit neuronal apoptosis induced by trophic factor deprivation.
CC   The anti-apoptotic tBAX proteins include tBAX70, tBAX78
CC   and their mutants. These proteins contain the N-terminal region and at
CC   least a portion of the BH3 domain of BAX alpha and lack the BH1, BH2 and
CC   C-terminal transmembrane domains. The tBAX protein lacking only the
CC   transmembrane domain has been shown to have anti-apoptotic activity.
CC   The tBAX proteins are used to treat diseases associated with neuronal
CC   apoptosis, e.g. neurodegenerative diseases, peripheral nerve injury,
CC   spinal cord injury, head trauma and stroke. tBAX proteins
CC   truncating at amino acid 68 of the BH3 domain of BAX alpha also have
CC   anti-apoptotic activity because Asp at position 68 is shown to be
CC   important for BAX activity and is conserved in all BCL-2 family members.
CC   The present sequence is a BH3 domain of mouse BAK protein, a
CC   pro-apoptotic protein belonging to BCL-2 family that is involved in
CC   regulation of neuronal programmed cell death. The present sequence
CC   contains the conserved Asp residue.
XX
SQ   Sequence   9 AA;

  Query Match             44.4%;  Score 20;  DB 21;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        2 LSGADLN 8
            | | |:|
Db        3 ligddin 9


Search completed: December 16, 2000, 03:07:18
Job time: 8059 sec
```

```
                    GenCore version 4.5
             Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 00:22:06 ; Search time 113.2 Seconds
                                             (without alignments)
                                             7.424 Million cell updates/sec

Title:          US-09-529-121-1
Perfect score:  45
Sequence:       1 YLSGANLNL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters:      297973

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      SPTREMBL_14:*
               1:  sp_archea:*
               2:  sp_bacteria:*
               3:  sp_fungi:*
               4:  sp_human:*
               5:  sp_invertebrate:*
               6:  sp_mammal:*
               7:  sp_mhc:*
               8:  sp_organelle:*
               9:  sp_phage:*
               10:  sp_plant:*
               11:  sp_rodent:*
               12:  sp_virus:*
               13:  sp_vertebrate:*
               14:  sp_unclassified:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                    %
Result              Query
  No.     Score  Match Length DB  ID                     Description
-----------------------------------------------------------------------------
    1       36    80.0    385  2  P94599                 P94599 bacillus th
    2       35    77.8    949  2  Q9RTF5                 Q9rtf5 deinococcus
    3       35    77.8    979  5  Q21962                 Q21962 caenorhabdi
    4       35    77.8    985  5  Q9VH09                 Q9vh09 drosophila
    5       35    77.8   1024 13  Q9PUU9                  Q9puu9 anas platyr
    6       34    75.6    131  8  Q35903                 Q35903 strongyloce
    7       34    75.6    259  2  Q55773                 Q55773 synechocyst
    8       34    75.6    272  5  Q9XVM8                 Q9xvm8 caenorhabdi
    9       34    75.6    554  2  O07411                 O07411 mycobacteri
   10       34    75.6    702  5  Q9VF80                 Q9vf80 drosophila
   11       34    75.6    952  2  O32915                 O32915 mycobacteri
   12       33    73.3    143  4  Q9UPC0                 Q9upc0 homo sapien
   13       33    73.3    180  4  Q14628                 Q14628 homo sapien
   14       33    73.3    356  5  O44592                 O44592 caenorhabdi
   15       33    73.3    473  2  Q9RB20                 Q9rb20 pectobacter
   16       33    73.3    507  2  O84150                 O84150 chlamydia t
   17       33    73.3    983  2  P74416                 P74416 synechocyst
   18       33    73.3   1031 10  O22575                  O22575 hordeum spo
   19       33    73.3   1032 10  Q38766                  Q38766 avena sativ
   20       33    73.3   1037 10  O82642                  O82642 arabidopsis
   21       33    73.3   1044 10  O80988                  O80988 arabidopsis
   22       33    73.3   1180  5  Q24163                 Q24163 drosophila
   23       33    73.3   1186  5  Q9VED3                 Q9ved3 drosophila
   24       33    73.3   1519  2  Q48237                 Q48237 helicobacte
   25       33    73.3   1985  4  Q9Y4D7                 Q9y4d7 homo sapien
   26       32    71.1    236  5  Q20421                 Q20421 caenorhabdi
   27       32    71.1    311  5  O01675                 O01675 branchiosto
   28       32    71.1    388  2  Q55201                 Q55201 synechocyst
   29       32    71.1    447 10  Q9SLV4                  Q9slv4 pisum sativ
   30       32    71.1    474  2  Q9WY57                 Q9wy57 thermotoga
   31       32    71.1    502  1  O57709                 O57709 pyrococcus
   32       32    71.1    502  1  Q9UXT1                 Q9uxt1 pyrococcus
   33       32    71.1    521  1  Q9YA18                 Q9ya18 aeropyrum p
   34       32    71.1    790  4  Q9Y2D7                 Q9y2d7 homo sapien
   35       32    71.1    858  3  P78988                 P78988 hebeloma ci
   36       32    71.1    877  5  Q9VT60                 Q9vt60 drosophila
   37       31    68.9    284  1  O58324                 O58324 pyrococcus
   38       31    68.9    297 12  Q64783                  Q64783 avian adeno
   39       31    68.9    298 12  Q9WA61                  Q9wa61 turnip mosa
   40       31    68.9    301  2  Q44404                 Q44404 agrobacteri
   41       31    68.9    324 10  O65678                  O65678 arabidopsis
   42       31    68.9    336 10  Q9ST82                  Q9st82 oryza sativ
   43       31    68.9    348  8  Q33154                 Q33154 suksdorfia
   44       31    68.9    351  8  Q31903                 Q31903 bolandra or
   45       31    68.9    351  8  Q31922                 Q31922 boykinia ro

                              ALIGNMENTS


RESULT   1
P94599
ID   P94599      PRELIMINARY;      PRT;   385 AA.
AC   P94599;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAY-1997 (TrEMBLrel. 03, Last annotation update)
DE   SERINE/THREONINE KINASE.
GN   PK1.
OS   Bacillus thuringiensis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=84-I-1-13;
RA   Dunn M.G., Ellar D.J.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Y10168; CAA71249.1; -.
SQ   SEQUENCE   385 AA;  43949 MW;  6013163292D0329E CRC64;

  Query Match             80.0%;  Score 36;  DB 2;  Length 385;
  Best Local Similarity   87.5%;  Pred. No. 18;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGANLN 8
            |||| |||
Db      278 YLSGPNLN 285


RESULT   2
Q9RTF5
ID   Q9RTF5      PRELIMINARY;      PRT;   949 AA.
AC   Q9RTF5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   GLYCINE CLEAVAGE SYSTEM P PROTEIN.
GN   DR1809.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE; 20036896.
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome Sequence of the Radioresistant Bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE002021; AAF11360.1; -.
DR   TIGR; DR1809; -.
SQ   SEQUENCE   949 AA;  102122 MW;  FDCBA42D4E0888D5 CRC64;

  Query Match             77.8%;  Score 35;  DB 2;  Length 949;
  Best Local Similarity   75.0%;  Pred. No. 78;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGANLN 8
            || |||:|
Db      668 YLDGANMN 675

RESULT   3
Q21962
ID   Q21962      PRELIMINARY;      PRT;   979 AA.
AC   Q21962;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   SIMILAR TO GLYCINE DEHYDROGENASE.
GN   R12C12.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Favello T.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U23510; AAC46780.1; -.
SQ   SEQUENCE   979 AA;  108865 MW;  F445CA0B46C7AA02 CRC64;

  Query Match             77.8%;  Score 35;  DB 5;  Length 979;
  Best Local Similarity   75.0%;  Pred. No. 81;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGANLN 8
            || |||:|
Db      702 YLDGANMN 709

RESULT   4
Q9VH09
ID   Q9VH09      PRELIMINARY;      PRT;   985 AA.
AC   Q9VH09;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG3999 PROTEIN.
GN   CG3999.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE; 20196006.
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003686; AAF54512.1; -.
DR   FLYBASE; FBgn0037801; CG3999.
SQ   SEQUENCE   985 AA;  109713 MW;  F8FBF5CFD251EDE7 CRC64;

  Query Match             77.8%;  Score 35;  DB 5;  Length 985;
  Best Local Similarity   75.0%;  Pred. No. 81;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGANLN 8
            || |||:|
Db      701 YLDGANMN 708
```

```
RESULT   5
Q9PUU9
ID   Q9PUU9      PRELIMINARY;      PRT;  1024 AA.
AC   Q9PUU9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   GLYCINE DECARBOXYLASE P PROTEIN.
OS   Anas platyrhynchos (Domestic duck).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anas.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEKIN DUCK; TISSUE=LIVER;
RX   MEDLINE; 96323119.
RA   Li J.S., Tong S.P., Wands J.R.;
RT   "Characterization of a 120-Kilodalton pre-S-binding protein as a
RT   candidate duck hepatitis B virus receptor.";
RL   J. Virol. 70:6029-6035(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEKIN DUCK; TISSUE=LIVER;
RA   Li J., Tong S., Wands J.R.;
RT   "Identification and Expression of Glycine Decarboxylase (p120) as a
RT   Duck Hepatitis B Virus Pre-S Envelope-binding Protein.";
RL   J. Biol. Chem. 274:27658-27665(1999).
DR   EMBL; AF137264; AAD56281.1; -.
SQ   SEQUENCE   1024 AA;  113473 MW;  657EFC89374DFEC3 CRC64;

  Query Match             77.8%;  Score 35;  DB 13;  Length 1024;
  Best Local Similarity   75.0%;  Pred. No. 85;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANLN 8
          || |||:|
Db    729 YLDGANMN 736

RESULT   6
Q35903
ID   Q35903      PRELIMINARY;      PRT;   131 AA.
AC   Q35903;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   NADH DEHYDROGENASE SUBUNIT 5 (ND5) (FRAGMENT).
OS   Strongylocentrotus pallidus.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
OC   Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91175684.
RA   Thomas W.K., Maa J., Wilson A.C.;
RT   "Shifting constraints on tRNA genes during mitochondrial DNA evolution
RT   in animals.";
RL   New Biol. 1:93-100(1989).
DR   EMBL; M27524; CAB25461.1; -.
KW   Mitochondrion.
FT   NON_TER     131    131
SQ   SEQUENCE   131 AA;  14405 MW;  903318CDAD6E4C0D CRC64;

  Query Match             75.6%;  Score 34;  DB 8;  Length 131;
  Best Local Similarity   66.7%;  Pred. No. 14;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANLNL 9
          ||| :|:||
```

```
Db     31 YLSNSNINL 39

RESULT   7
Q55773
ID   Q55773      PRELIMINARY;      PRT;   259 AA.
AC   Q55773;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HYPOTHETICAL 27.4 KDA PROTEIN.
GN   SLL0183.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RA   Tabata S.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC6803;
RX   MEDLINE; 96127529.
RA   Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA   Sugiura M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT   region from map positions 64% to 92% of the genome.";
RL   DNA Res. 2:153-166(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97061201.
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D64002; BAA10422.1; -.
DR   INTERPRO; IPR001440; -.
DR   INTERPRO; IPR001646; -.
DR   PFAM; PF00805; Pentapeptide; 2.
KW   Hypothetical protein.
SQ   SEQUENCE   259 AA;  27402 MW;  B6D54CA59C6FCDE7 CRC64;

  Query Match             75.6%;  Score 34;  DB 2;  Length 259;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 LSGANLN 8
          |||||||
Db     91 LSGANLN 97

RESULT   8
Q9XVM8
ID   Q9XVM8      PRELIMINARY;      PRT;   272 AA.
AC   Q9XVM8;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   F53F1.6 PROTEIN.
GN   F53F1.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
```

```
RA   Burton J.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z81088; CAB03123.1; -.
SQ   SEQUENCE   272 AA;  31908 MW;  B00BDE57AFD40584 CRC64;

  Query Match             75.6%;  Score 34;  DB 5;  Length 272;
  Best Local Similarity   66.7%;  Pred. No. 31;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGANLNL 9
                || | |||:
Db        235 YLEGVNLNV 243

RESULT   9
O07411
ID   O07411      PRELIMINARY;      PRT;   554 AA.
AC   O07411;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   PROBABLE FATTY-ACID COA LIGASE.
GN   RV0166 OR MTCI28.06.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 98295987.
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Sqares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
DR   EMBL; Z97050; CAB09749.1; -.
DR   HSSP; P08659; 1LCI.
DR   TUBERCULIST; Rv0166; -.
DR   INTERPRO; IPR000873; -.
DR   PFAM; PF00501; AMP-binding; 1.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Ligase.
SQ   SEQUENCE   554 AA;  59905 MW;  3AC3100FAF0B9E88 CRC64;

  Query Match             75.6%;  Score 34;  DB 2;  Length 554;
  Best Local Similarity   75.0%;  Pred. No. 69;
```

```
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGANLN 8
              | ||||:|
Db        217 YTSGANIN 224


RESULT  10
Q9VF80
ID   Q9VF80      PRELIMINARY;      PRT;   702 AA.
AC   Q9VF80;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update).
DE   CG6194 PROTEIN.
GN   CG6194.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE; 20196006.
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003708; AAF55180.1; -.
DR   FLYBASE; FBgn0038325; CG6194.
SQ   SEQUENCE   702 AA;  78620 MW;  0A3222A1886CFEE2 CRC64;

  Query Match             75.6%;  Score 34;  DB 5;  Length 702;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 SGANLNL 9
              |||||||
```

```
Db        114 SGANLNL 120


RESULT  11
O32915
ID   O32915      PRELIMINARY;      PRT;   952 AA.
AC   O32915;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   GLYCINE DEHYDROGENASE (DECARBOXYLATING).
GN   GCVP.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Skelton J., Churcher C.M.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93188700.
RA   Eiglmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT   "Use of an ordered cosmid library to deduce the genomic organization
RT   of Mycobacterium leprae.";
RL   Mol. Microbiol. 7:197-206(1993).
DR   EMBL; AL008609; CAA15464.1; -.
SQ   SEQUENCE   952 AA;  101282 MW;  ADC68B6265C3D8FF CRC64;


  Query Match             75.6%;  Score 34;  DB 2;  Length 952;
  Best Local Similarity   75.0%;  Pred. No. 1.3e+02;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGANLN 8
              |: |||||
Db        674 YVDGANLN 681


RESULT  12
Q9UPC0
ID   Q9UPC0      PRELIMINARY;      PRT;   143 AA.
AC   Q9UPC0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   INTERLEUKIN-1 INTRACELLULAR RECEPTOR ANTAGONIST VARIANT (FRAGMENT).
GN   IL1RN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98183404.
RA   Weissbach L., Tran K., Colquhoun S.A., Champliaud M.F., Towle C.A.;
RT   "Detection of an interleukin-1 intracellular receptor antagonist mRNA
RT   variant.";
RL   Biochem. Biophys. Res. Commun. 244:91-95(1998).
DR   EMBL; AF043143; AAC39672.1; -.
DR   INTERPRO; IPR000975; -.
DR   PFAM; PF00340; interleukin-1; 1.
DR   PRINTS; PR00264; INTERLEUKIN1.
DR   PROSITE; PS00253; INTERLEUKIN_1; 1.
KW   Receptor.
FT   NON_TER     143    143
SQ   SEQUENCE   143 AA;  16142 MW;  4CAD6784B890906B CRC64;


  Query Match             73.3%;  Score 33;  DB 4;  Length 143;
  Best Local Similarity   66.7%;  Pred. No. 25;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGANLNL 9
              || | |:||
Db         25 YLQGPNVNL 33


RESULT  13
Q14628
ID   Q14628      PRELIMINARY;      PRT;   180 AA.
AC   Q14628;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   INTRACELLULAR IL-1 RECEPTOR ANTAGONIST TYPE II.
GN   IL-1RN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95355865.
RA   Muzio M., Polentarutti N., Sironi M., Poli G., De Gioia L.,
RA   Introna M., Mantovani A., Colotta F.;
RT   "Cloning and characterization of a new isoform of the interleukin 1
RT   receptor antagonist.";
RL   J. Exp. Med. 182:623-628(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97146044.
RA   Jenkins J.K., Drong R.F., Shuck M.E., Bienkowski M.J., Slightom J.L.,
RA   Arend W.P., Smith M.F. Jr.;
RT   "Intracellular IL-1 receptor antagonist promoter: cell type-specific
RT   and inducible regulatory regions.";
RL   J. Immunol. 158:748-755(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Slightom J.L.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X84348; CAA59087.1; -.
DR   EMBL; U65590; AAB92269.1; -.
DR   HSSP; P18510; 1IRP.
DR   INTERPRO; IPR000975; -.
DR   PFAM; PF00340; interleukin-1; 1.
DR   PRINTS; PR00264; INTERLEUKIN1.
DR   PROSITE; PS00253; INTERLEUKIN_1; 1.
SQ   SEQUENCE   180 AA;  19897 MW;  624A1574C2334229 CRC64;


  Query Match             73.3%;  Score 33;  DB 4;  Length 180;
  Best Local Similarity   66.7%;  Pred. No. 32;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSGANLNL 9
              || | |:||
Db         62 YLQGPNVNL 70


RESULT  14
O44592
ID   O44592      PRELIMINARY;      PRT;   356 AA.
AC   O44592;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   F48G7.1 PROTEIN.
GN   F48G7.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Clarke K., Wohldmann P., Harrison M.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF039044; AAC47951.1; -.
DR   INTERPRO; IPR000168; -.
DR   INTERPRO; IPR003002; -.
DR   PFAM; PF01461; 7tm_4; 1.
SQ   SEQUENCE   356 AA;  41323 MW;  DEF906750193F6A4 CRC64;

  Query Match             73.3%;  Score 33;  DB 5;  Length 356;
  Best Local Similarity   66.7%;  Pred. No. 68;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSGANLNL 9
            || |||| :
Db      152 YLGGANLEI 160


RESULT 15
Q9RB20
ID   Q9RB20      PRELIMINARY;      PRT;   473 AA.
AC   Q9RB20;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   METALLOPROTEASE.
GN   PRTW.
OS   Pectobacterium carotovorum subsp. cartovorum.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Erwinia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SCC3193;
RX   MEDLINE; 99392458.
RA   Marits R., Koiv V., Laasik E., Mae A.;
RT   "Isolation of an extracellular protease gene of Erwinia carotovora
RT   subsp. carotovora strain SCC3193 by transposon mutagenesis and the
RT   role of protease in phytopathogenicity.";
RL   Microbiology 145:1959-1966(1999).
DR   EMBL; AF141295; AAD49575.1; -.
DR   INTERPRO; IPR000130; -.
DR   INTERPRO; IPR001343; -.
DR   PFAM; PF00353; hemolysinCabind; 1.
DR   PRINTS; PR00313; CABNDNGRPT.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR   PROSITE; PS00330; HEMOLYSIN_CALCIUM; 1.
KW   Protease; Metalloprotease.
SQ   SEQUENCE   473 AA;  50959 MW;  419522745CCBA373 CRC64;

  Query Match             73.3%;  Score 33;  DB 2;  Length 473;
  Best Local Similarity   75.0%;  Pred. No. 93;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGANLN 8
            || |||:|
Db      250 YLYGANMN 257


Search completed: December 16, 2000, 02:00:56
Job time: 5930 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000  Compugen Ltd.


OM protein - protein search, using sw model

Run on:          December 16, 2000, 00:51:20 ; Search time 69.45 Seconds
                                                (without alignments)
                                                4.140 Million cell updates/sec

Title:           US-09-529-121-1
Perfect score:   45
Sequence:        1 YLSGANLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters:      87993

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_39:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                   %
Result            Query
  No.    Score    Match Length DB  ID                         Description
---------------------------------------------------------------------------
    1       45    100.0    702  1  CCEM_HUMAN                 P06731 homo sapien
    2       35     77.8    956  1  GCSP_ECOLI                 P33195 escherichia
    3       35     77.8   1004  1  GCSP_CHICK                 P15505 gallus gall
    4       35     77.8   1020  1  GCSP_HUMAN                 P23378 homo sapien
    5       35     77.8   1034  1  GCSP_YEAST                 P49095 saccharomyc
    6       34     75.6    513  1  THD1_HAEIN                 P46493 haemophilus
    7       34     75.6    941  1  GCSP_MYCTU                 Q50601 mycobacteri
    8       33     73.3    177  1  IL1X_HUMAN                 P18510 homo sapien
    9       33     73.3   1034  1  GCSB_FLAPR                 P49362 flaveria pr
   10       33     73.3   1034  1  GCSP_FLAAN                 O49850 flaveria an
   11       33     73.3   1034  1  GCSP_FLATR                 O49852 flaveria tr
   12       33     73.3   1035  1  GCSP_SOLTU                 O49954 solanum tub
   13       33     73.3   1037  1  GCSA_FLAPR                 P49361 flaveria pr
   14       33     73.3   1057  1  GCSP_PEA                   P26969 pisum sativ
   15       32     71.1    335  1  YJ89_PYRHO                 O57713 pyrococcus
   16       32     71.1    350  1  PRIM_METJA                 Q58249 methanococc
   17       32     71.1    455  1  NTPA_PEA                   P52914 pisum sativ
   18       32     71.1    488  1  GCS2_BACSU                 P54377 bacillus su
   19       32     71.1    576  1  THDH_YEAST                 P00927 saccharomyc
   20       32     71.1    641  1  FIB2_PETMA                 P33573 petromyzon
   21       32     71.1    828  1  PMFC_PROMI                 P53514 proteus mir
   22       31     68.9    108  1  HLYU_VIBCH                 P52695 vibrio chol
   23       31     68.9    109  1  PER_SYRPI                  Q26612 syritta pip
   24       31     68.9    313  1  SURA_HAEIN                 P44721 haemophilus
   25       31     68.9    349  1  CGM6_HUMAN                 P31997 homo sapien
   26       31     68.9    357  1  YFJN_ECOLI                 P52129 escherichia
   27       31     68.9    445  1  YM85_MYCTU                 Q50680 mycobacteri
   28       31     68.9    514  1  THD1_ECOLI                 P04968 escherichia
   29       31     68.9    514  1  THD1_SALTY                 P20506 salmonella
   30       31     68.9    547  1  FIB1_ADE40                 P18047 human adeno
   31       31     68.9    562  1  FIB1_ADE41                 P14267 human adeno
   32       31     68.9    575  1  HEX3_ADEG1                 Q64754 avian adeno
   33       31     68.9    741  1  RN5A_HUMAN                 Q05823 homo sapien
   34       31     68.9    769  1  YCYL_CAEEL                 Q22523 caenorhabdi
   35       31     68.9    933  1  ODO1_ECOLI                 P07015 escherichia
   36       31     68.9   1017  1  GCSP_SCHPO                 Q09785 schizosacch
   37       31     68.9   1167  1  CLAA_BACTU                 P56956 bacillus th
   38       31     68.9   1286  1  AIDA_ECOLI                 Q03155 escherichia
   39       30     66.7    122  1  YNR1_YEAST                 P53887 saccharomyc
   40       30     66.7    156  1  AROQ_STRCO                 P15474 streptomyce
   41       30     66.7    246  1  PHBB_CHRVI                 P45375 chromatium
   42       30     66.7    270  1  PMIP_NICAL                 P49173 nicotiana a
   43       30     66.7    426  1  PSGB_HUMAN                 Q00887 homo sapien
   44       30     66.7    440  1  YGR4_YEAST                 P53107 saccharomyc
   45       30     66.7    510  1  YDYD_SCHPO                 O13692 schizosacch


                                  ALIGNMENTS


RESULT   1
CCEM_HUMAN
ID   CCEM_HUMAN     STANDARD;      PRT;   702 AA.
AC   P06731;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINOEMBRYONIC ANTIGEN PRECURSOR (CEA) (MECONIUM ANTIGEN 100)
DE   (CD66E ANTIGEN).
GN   CEA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90258861.
RA   Schrewe H., Thompson J., Bona M., Hefta L.J.F., Maruya A.,
RA   Hassauer M., Shively J.E., von Kleist S., Zimmermann W.;
RT   "Cloning of the complete gene for carcinoembryonic antigen: analysis
RT   of its promoter indicates a region conveying cell type-specific
RT   expression.";
RL   Mol. Cell. Biol. 10:2738-2748(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88038876.
RA   Beauchemin N., Benchimol S., Cournoyer D., Fuks A., Stanners C.P.; \
RT   "Isolation and characterization of full-length functional cDNA clones
RT   for human carcinoembryonic antigen.";
RL   Mol. Cell. Biol. 7:3221-3230(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89122014.
RA   Barnett T., Goebel S.J., Nothdurft M.A., Elting J.J.;
RT   "Carcinoembryonic antigen family: characterization of cDNAs coding
RT   for NCA and CEA and suggestion of nonrandom sequence variation in
RT   their conserved loop-domains.";
RL   Genomics 3:59-66(1988).
RN   [4]
RP   SEQUENCE OF 5-702 FROM N.A.
RX   MEDLINE; 87128144.
RA   Oikawa S., Nakazato H., Kosaki G.;
RT   "Primary structure of human carcinoembryonic antigen (CEA) deduced
RT   from cDNA sequence.";
RL   Biochem. Biophys. Res. Commun. 142:511-518(1987).
RN   [5]
RP   SEQUENCE OF 331-702 FROM N.A.
RX   MEDLINE; 87204247.
RA   Zimmermann W., Ortlieb B., Friedrich R., von Kleist S.;
RT   "Isolation and characterization of cDNA clones encoding the human
RT   carcinoembryonic antigen reveal a highly conserved repeating
RT   structure.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:2960-2964(1987).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC   -!- TISSUE SPECIFICITY: FOUND IN ADENOCARCINOMAS OF ENDODERMALLY
CC       DERIVED DIGESTIVE SYSTEM EPITHELIUM AND FETAL COLON.
CC   -!- PTM: COMPLEX IMMUNOREACTIVE GLYCOPROTEIN WITH A MW OF 180 KDA
```

```
CC       COMPRISING 60% CARBOHYDRATE.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       7 IG-LIKE DOMAINS. BELONGS TO THE CARCINOEMBRYONIC ANTIGEN
CC       SUBFAMILY.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD66e entry;
CC       WWW="HTTP://WWW.NCBI.NLM.NIH.GOV/PROW/CD/CD66E.HTM".
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M17303; AAB59513.1; -.
DR   EMBL; M59262; AAA62835.1; ALT_SEQ.
DR   EMBL; M59255; AAA62835.1; JOINED.
DR   EMBL; M59256; AAA62835.1; JOINED.
DR   EMBL; M59257; AAA62835.1; JOINED.
DR   EMBL; M59258; AAA62835.1; JOINED.
DR   EMBL; M59259; AAA62835.1; JOINED.
DR   EMBL; M59260; AAA62835.1; JOINED.
DR   EMBL; M59261; AAA62835.1; JOINED.
DR   EMBL; M59709; -; NOT_ANNOTATED_CDS.
DR   EMBL; M59710; -; NOT_ANNOTATED_CDS.
DR   EMBL; M29540; AAA51967.1; -.
DR   EMBL; X16455; CAA34474.1; -.
DR   EMBL; M15042; AAA51963.1; -.
DR   EMBL; M16234; AAA51972.1; -.
DR   PIR; A36319; A36319.
DR   MIM; 114890; -.
DR   INTERPRO; IPR003006; -.
DR   PFAM; PF00047; ig; 7.
KW   Immunoglobulin domain; Glycoprotein; GPI-anchor; Membrane; Signal.
FT   SIGNAL        1     34
FT   CHAIN        35      ?       CARCINOEMBRYONIC ANTIGEN.
FT   PROPEP        ?    702       REMOVED IN MATURE FORM.
FT   DOMAIN       35    144       IG-LIKE DOMAIN 1.
FT   DOMAIN      146    237       IG-LIKE DOMAIN 2.
FT   DOMAIN      238    322       IG-LIKE DOMAIN 3.
FT   DOMAIN      324    415       IG-LIKE DOMAIN 4.
FT   DOMAIN      416    498       IG-LIKE DOMAIN 5.
FT   DOMAIN      502    593       IG-LIKE DOMAIN 6.
FT   DOMAIN      594    677       IG-LIKE DOMAIN 7.
FT   CARBOHYD    104    104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    115    115       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    152    152       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    182    182       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    197    197       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    204    204       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    208    208       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    246    246       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    256    256       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    274    274       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    288    288       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    292    292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    309    309       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    330    330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    351    351       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    360    360       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    375    375       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    432    432       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    466    466       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    480    480       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    508    508       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    529    529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    553    553       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    560    560       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    580    580       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    612    612       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    650    650       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    665    665       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CONFLICT    320    320       MISSING (IN REF. 4).
SQ   SEQUENCE   702 AA;  76795 MW;  6299AE26CDDBDB5C CRC64;

  Query Match             100.0%;  Score 45;  DB 1;  Length 702;
  Best Local Similarity   100.0%;  Pred. No. 0.14;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGANLNL 9
              |||||||||
Db        605 YLSGANLNL 613

RESULT   2
GCSP_ECOLI
ID   GCSP_ECOLI     STANDARD;      PRT;   956 AA.
AC   P33195;
DT   01-OCT-1993 (Rel. 27, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GLYCINE DEHYDROGENASE [DECARBOXYLATING] (EC 1.4.4.2) (GLYCINE
DE   DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-PROTEIN).
GN   GCVP.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=K12;
RX   MEDLINE; 94237484.
RA   Stauffer L.T., Fogarty S.J., Stauffer G.V.;
RT   "Characterization of the Escherichia coli gcv operon.";
RL   Gene 142:17-22(1994).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC   STRAIN=K12 / W3110;
RX   MEDLINE; 93387305.
RA   Okamura-Ikeda K., Ohmura Y., Fujiwara K., Motokawa Y.;
RT   "Cloning and nucleotide sequence of the gcv operon encoding the
RT   Escherichia coli glycine-cleavage system.";
RL   Eur. J. Biochem. 216:539-548(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC       GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE
CC       THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND
CC       THE REMAINING METHYLAMINE MOIETY IS THE THEN TRANSFERRED TO THE
CC       LIPOAMIDE COFACTOR OF THE H PROTEIN.
CC   -!- CATALYTIC ACTIVITY: GLYCINE + LIPOYLPROTEIN = S-AMINOMETHYL-
CC       DIHYDROLIPOYLPROTEIN + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- SUBUNIT: THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF FOUR PROTEINS:
CC       P, T, L, AND H.
CC   -!- INDUCTION: BY GLYCINE.
CC   -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L20872; AAA23867.1; -.
```

```
DR   EMBL; X73958; CAA52146.1; -.
DR   EMBL; U28377; AAA69071.1; -.
DR   EMBL; AE000373; AAC75941.1; -.
DR   PIR; S36834; S36834.
DR   ECOGENE; EG11810; GCVP.
KW   Oxidoreductase; Pyridoxal phosphate.
FT   INIT_MET      0      0
FT   BINDING     707    707       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   956 AA;  104245 MW;  28B7A78D1D0DB6F7 CRC64;


  Query Match             77.8%;  Score 35;  DB 1;  Length 956;
  Best Local Similarity   75.0%;  Pred. No. 21;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGANLN 8
            || |||:|
Db      678 YLDGANMN 685


RESULT   3
GCSP_CHICK
ID   GCSP_CHICK     STANDARD;      PRT;  1004 AA.
AC   P15505;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLYCINE DEHYDROGENASE [DECARBOXYLATING], MITOCHONDRIAL PRECURSOR
DE   (EC 1.4.4.2) (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-
DE   PROTEIN).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91131643.
RA   Kume A., Koyata H., Sakakibara T., Ishiguro Y., Kure S., Hiraga K.;
RT   "The glycine cleavage system. Molecular cloning of the chicken and
RT   human glycine decarboxylase cDNAs and some characteristics involved
RT   in the deduced protein structures.";
RL   J. Biol. Chem. 266:3323-3329(1991).
RN   [2]
RP   SEQUENCE OF 704-757.
RC   TISSUE=LIVER;
RX   MEDLINE; 88106483.
RA   Fujiwara K., Okamura-Ikeda K., Motokawa Y.;
RT   "Amino acid sequence of the phosphopyridoxyl peptide from P-protein
RT   of the chicken liver glycine cleavage system.";
RL   Biochem. Biophys. Res. Commun. 149:621-627(1987).
CC   -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC       GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE
CC       THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND
CC       THE REMAINING METHYLAMINE MOIETY IS THE THEN TRANSFERRED TO THE
CC       LIPOAMIDE COFACTOR OF THE H PROTEIN.
CC   -!- CATALYTIC ACTIVITY: GLYCINE + LIPOYLPROTEIN = S-AMINOMETHYL-
CC       DIHYDROLIPOYLPROTEIN + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- SUBUNIT: HOMODIMER. THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF
CC       FOUR PROTEINS: P, T, L, AND H.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC   -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M64402; AAA49029.1; -.
DR   EMBL; D90266; BAA14313.1; -.
```

```
DR   EMBL; D90240; BAA14287.1; -.
DR   PIR; A27483; A27483.
DR   PIR; A39521; A39521.
KW   Oxidoreductase; Mitochondrion; Pyridoxal phosphate; Transit peptide.
FT   TRANSIT       1      ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN         ?   1004       GLYCINE DEHYDROGENASE [DECARBOXYLATING].
FT   BINDING     738    738       PYRIDOXAL PHOSPHATE.
SQ   SEQUENCE   1004 AA;  111852 MW;  4446D7C66E0DC4BD CRC64;

  Query Match             77.8%;  Score 35;  DB 1;  Length 1004;
  Best Local Similarity   75.0%;  Pred. No. 22;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGANLN 8
            || |||:|
Db      709 YLDGANMN 716


RESULT   4
GCSP_HUMAN
ID   GCSP_HUMAN     STANDARD;      PRT;  1020 AA.
AC   P23378;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   GLYCINE DEHYDROGENASE [DECARBOXYLATING], MITOCHONDRIAL PRECURSOR
DE   (EC 1.4.4.2) (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-
DE   PROTEIN).
GN   GCSP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91144593.
RA   Kure S., Narisawa K., Tada K.;
RT   "Structural and expression analyses of normal and mutant mRNA
RT   encoding glycine decarboxylase: three-base deletion in mRNA causes
RT   nonketotic hyperglycinemia.";
RL   Biochem. Biophys. Res. Commun. 174:1176-1182(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91131643.
RA   Kume A., Koyata H., Sakakibara T., Ishiguro Y., Kure S., Hiraga K.;
RT   "The glycine cleavage system. Molecular cloning of the chicken and
RT   human glycine decarboxylase cDNAs and some characteristics involved
RT   in the deduced protein structures.";
RL   J. Biol. Chem. 266:3323-3329(1991).
RN   [3]
RP   VARIANT NKH1 ILE-564.
RX   MEDLINE; 92340654.
RA   Kure S., Takayanagi M., Narisawa K., Tada K., Leisti J.;
RT   "Identification of a common mutation in Finnish patients with
RT   nonketotic hyperglycinemia.";
RL   J. Clin. Invest. 90:160-164(1992).
CC   -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC       GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE
CC       THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND
CC       THE REMAINING METHYLAMINE MOIETY IS THE THEN TRANSFERRED TO THE
CC       LIPOAMIDE COFACTOR OF THE H PROTEIN.
CC   -!- CATALYTIC ACTIVITY: GLYCINE + LIPOYLPROTEIN = S-AMINOMETHYL-
CC       DIHYDROLIPOYLPROTEIN + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- SUBUNIT: HOMODIMER. THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF
CC       FOUR PROTEINS: P, T, L, AND H.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC   -!- DISEASE: DEFECTS IN GCSP ARE THE CAUSE OF TYPE I NONKETOTIC
CC       HYPERGLYCINEMIA (NKH1). NKH IS AN AUTOSOMAL RECESSIVE DISEASE
CC       CHARACTERIZED BY ACCUMULATION OF A LARGE AMOUNT OF GLYCINE IN BODY
CC       FLUID AND BY SEVERE NEUROLOGICAL SYMPTOMES.
CC   -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.
CC   --------------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M63635; AAA36478.1; -.
DR   EMBL; M64590; AAA36463.1; -.
DR   EMBL; D90239; BAA14286.1; -.
DR   PIR; B39521; B39521.
DR   PIR; JN0124; JN0124.
DR   MIM; 238300; -.
KW   Oxidoreductase; Mitochondrion; Pyridoxal phosphate; Transit peptide;
KW   Disease mutation.
FT   TRANSIT       1     35       MITOCHONDRION (POTENTIAL).
FT   CHAIN        36   1020       GLYCINE DEHYDROGENASE [DECARBOXYLATING].
FT   BINDING     754    754       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   VARIANT     564    564       S -> I (IN NKH1).
FT                                /FTId=VAR_004979.
FT   CONFLICT    396    396       A -> R (IN REF. 2).
FT   CONFLICT    608    608       H -> Y (IN REF. 2).
FT   CONFLICT    976    976       V -> M (IN REF. 2).
SQ   SEQUENCE   1020 AA;  112712 MW;  129FAF26B8D37E0F CRC64;

  Query Match             77.8%;  Score 35;  DB 1;  Length 1020;
  Best Local Similarity   75.0%;  Pred. No. 23;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANLN 8
          || |||:|
Db    725 YLDGANMN 732

RESULT   5
GCSP_YEAST
ID   GCSP_YEAST     STANDARD;      PRT;  1034 AA.
AC   P49095;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLYCINE DEHYDROGENASE [DECARBOXYLATING], MITOCHONDRIAL PRECURSOR
DE   (EC 1.4.4.2) (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-
DE   PROTEIN).
GN   GCV2 OR GSD2 OR YMR189W OR YM9646.01.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96120340.
RA   Sinclair D.A., Dawes I.W.;
RT   "Genetics of the synthesis of serine from glycine and the utilization
RT   of glycine as sole nitrogen source by Saccharomyces cerevisiae.";
RL   Genetics 140:1213-1222(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Pearson D., Bowman S., Barrell B.G., Rajandream M.A.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC       GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE
CC       THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND
CC       THE REMAINING METHYLAMINE MOIETY IS THE THEN TRANSFERRED TO THE
CC       LIPOAMIDE COFACTOR OF THE H PROTEIN.
CC   -!- CATALYTIC ACTIVITY: GLYCINE + LIPOYLPROTEIN = S-AMINOMETHYL-
CC       DIHYDROLIPOYLPROTEIN + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL (BY SIMILARITY).
CC   -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.
CC   --------------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U20641; AAB18933.1; -.
DR   EMBL; Z47815; CAA87810.1; -.
DR   SGD; S0004801; GCV2.
KW   Oxidoreductase; Mitochondrion; Pyridoxal phosphate; Transit peptide.
FT   TRANSIT       1      ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN         ?   1034       GLYCINE DEHYDROGENASE [DECARBOXYLATING].
FT   BINDING     773    773       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   1034 AA;  114451 MW;  F4D52642B0BDA041 CRC64;

  Query Match             77.8%;  Score 35;  DB 1;  Length 1034;
  Best Local Similarity   75.0%;  Pred. No. 23;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANLN 8
          || |||:|
Db    744 YLDGANMN 751

RESULT   6
THD1_HAEIN
ID   THD1_HAEIN     STANDARD;      PRT;   513 AA.
AC   P46493;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   THREONINE DEHYDRATASE BIOSYNTHETIC (EC 4.2.1.16) (THREONINE
DE   DEAMINASE).
GN   ILVA OR HI0738.1.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   IDENTIFICATION.
RA   Koonin E.V., Rudd K.E.;
RL   Submitted (SEP-1995) to the SWISS-PROT data bank.
CC   -!- FUNCTION: DEAMINATES L-THREONINE, BUT ALSO L-SERINE TO A LESSER
CC       EXTENT (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: L-THREONINE + H(2)O = 2-OXOBUTANOATE + NH(3)
CC       + H(2)O.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC   -!- ENZYME REGULATION: ISOLEUCINE ALLOSTERICALLY INHIBITS WHILE
CC       VALINE ALLOSTERICALLY ACTIVATES THIS ENZYME (BY SIMILARITY).
CC   -!- PATHWAY: FIRST STEP OF ISOLEUCINE BIOSYNTHESIS.
CC   -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC   -!- SIMILARITY: TO THREONINE DEHYDRATASE CATABOLIC AND SERINE
CC       DEHYDRATASE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U32757; AAC22398.1; -.
DR   TIGR; HI0738.1; -.
DR   INTERPRO; IPR000634; -.
DR   INTERPRO; IPR001721; -.
DR   INTERPRO; IPR001926; -.
DR   PFAM; PF00291; S_T_dehydratase; 1.
DR   PFAM; PF00585; Thr_dehydrat_C; 2.
DR   PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW   Isoleucine biosynthesis; Lyase; Pyridoxal phosphate;
KW   Allosteric enzyme.
FT   BINDING      63     63       PYRIDOXAL PHOSPHATE.
SQ   SEQUENCE   513 AA;  56662 MW;  DF42CA8B6FDE4CD7 CRC64;


  Query Match             75.6%;  Score 34;  DB 1;  Length 513;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 LSGANLN 8
          |||||||
Db    315 LSGANLN 321


RESULT   7
GCSP_MYCTU
ID   GCSP_MYCTU     STANDARD;      PRT;   941 AA.
AC   Q50601;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROBABLE GLYCINE DEHYDROGENASE [DECARBOXYLATING] (EC 1.4.4.2) (GLYCINE
DE   DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-PROTEIN).
GN   GCVP OR GCVB OR RV1832 OR MTCY1A11.11C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 98295987.
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Sqares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC       GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE
CC       THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND
CC       THE REMAINING METHYLAMINE MOIETY IS THE THEN TRANSFERRED TO THE
CC       LIPOAMIDE COFACTOR OF THE H PROTEIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: GLYCINE + LIPOYLPROTEIN = S-AMINOMETHYL-
CC       DIHYDROLIPOYLPROTEIN + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC   -!- SUBUNIT: THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF FOUR PROTEINS:
CC       P, T, L, AND H (BY SIMILARITY).
CC   -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z78020; CAB01470.1; -.
DR   TUBERCULIST; RV1832; -.
KW   Oxidoreductase; Pyridoxal phosphate.
FT   DOMAIN      437    446       ALA-RICH.
FT   BINDING     692    692       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   941 AA;  99510 MW;  BFFDE54EDA56B914 CRC64;


  Query Match             75.6%;  Score 34;  DB 1;  Length 941;
  Best Local Similarity   75.0%;  Pred. No. 33;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANLN 8
           |: |||||
Db    663 YVDGANLN 670


RESULT   8
IL1X_HUMAN
ID   IL1X_HUMAN     STANDARD;      PRT;   177 AA.
AC   P18510;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   INTERLEUKIN-1 RECEPTOR ANTAGONIST PROTEIN PRECURSOR (IL-1RA) (ICIL-
DE   1RA) (IRAP) (IL-1RN).
GN   IL1RN OR IL1RA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90220867.
RA   Carter D.B., Deibel M.R. Jr., Dunn C.J., Tomich C.S.C., Laborde A.L.,
RA   Slightom J.L., Berger A.E., Bienkowski M.J., Sun F.F., McEwan R.N.,
RA   Harris P.K.W., Yem A.W., Waszak G.A., Chosay J.G., Sieu L.C.,
RA   Hardee M.M., Zurcher-Neely H.A., Reardon I.M., Heinrikson R.L.,
RA   Truesdell S.E., Shelly J.A., Eessalu T.E., Taylor B.M., Tracey D.E.;
RT   "Purification, cloning, expression and biological characterization of
RT   an interleukin-1 receptor antagonist protein.";
RL   Nature 344:633-638(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90136921.
RA   Eisenberg S.P., Evans R.J., Arend W.P., Verderber E., Brewer M.T.,
RA   Hannum C.H., Thompson R.C.;
RT   "Primary structure and functional expression from complementary DNA
RT   of a human interleukin-1 receptor antagonist.";
RL   Nature 343:341-346(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91271363.
RA   Eisenberg S.P., Brewer M.T., Verderber E., Heimdal P.,
RA   Brandhuber B.J., Thompson R.C.;
RT   "Interleukin 1 receptor antagonist is a member of the interleukin 1
RT   gene family: evolution of a cytokine control mechanism.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:5232-5236(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92338323.
RA   Lennard A., Gorman P., Carrier M., Griffiths S., Scotney H.,
RA   Sheer D., Solari R.;
RT   "Cloning and chromosome mapping of the human interleukin-1 receptor
RT   antagonist gene.";
RL   Cytokine 4:83-89(1992).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97146044.
```

```
RA   Jenkins J.K., Drong R.F., Shuck M.E., Bienkowski M.J., Slightom J.L.,
RA   Arend W.P., Smith M.F. Jr.;
RT   "Intracellular IL-1 receptor antagonist promoter: cell type-specific
RT   and inducible regulatory regions.";
RL   J. Immunol. 158:748-755(1997).
RN   [6]
RP   SEQUENCE OF 26-45.
RX   MEDLINE; 90136920.
RA   Hannum C.H., Wilcox C.J., Arend W.P., Joslin F.G., Dripps D.J.,
RA   Heimdal P.L., Armes L.G., Sommer A., Eisenberg S.P., Thompson R.C.;
RT   "Interleukin-1 receptor antagonist activity of a human interleukin-1
RT   inhibitor.";
RL   Nature 343:336-340(1990).
RN   [7]
RP   SEQUENCE OF 26-52.
RX   MEDLINE; 90354444.
RA   Bienkowski M.J., Eessalu T.E., Berger A.E., Truesdell S.E.,
RA   Shelly J.A., Laborde A.L., Zurcher-Neely H.A., Reardon I.M.,
RA   Heinrikson R.L., Chosay J.G., Tracey D.E.;
RT   "Purification and characterization of interleukin 1 receptor level
RT   antagonist proteins from THP-1 cells.";
RL   J. Biol. Chem. 265:14505-14511(1990).
RN   [8]
RP   SEQUENCE FROM N.A. (INTRACELLULAR FORM).
RX   MEDLINE; 91219436.
RA   Haskill S., Martin G., van Le L., Morris J., Peace A., Bigler C.F.,
RA   Jaffe G.J., Hammerberg C., Sporn S.A., Fong S., Arend W.P., Ralph P.;
RT   "cDNA cloning of an intracellular form of the human interleukin 1
RT   receptor antagonist associated with epithelium.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:3681-3685(1991).
RN   [9]
RP   STRUCTURE BY NMR.
RX   MEDLINE; 92297633.
RA   Stockman B.J., Scahill T.A., Roy M., Ulrich E.L., Strakalaitis N.A.,
RA   Brunner D.P., Yem A.W., Deibel M.R. Jr.;
RT   "Secondary structure and topology of interleukin-1 receptor
RT   antagonist protein determined by heteronuclear three-dimensional NMR
RT   spectroscopy.";
RL   Biochemistry 31:5237-5244(1992).
RN   [10]
RP   STRUCTURE BY NMR.
RX   MEDLINE; 94320651.
RA   Stockman B.J., Scahill T.A., Strakalaitis N.A., Brunner D.P.,
RA   Yem A.W., Deibel M.R. Jr.;
RT   "Solution structure of human interleukin-1 receptor antagonist
RT   protein.";
RL   FEBS Lett. 349:79-83(1994).
RN   [11]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX   MEDLINE; 94230368.
RA   Vigers G.P.A., Caffes P., Evans R.J., Thompson R.C., Eisenberg S.P.,
RA   Brandhuber B.J.;
RT   "X-ray structure of interleukin-1 receptor antagonist at 2.0-A
RT   resolution.";
RL   J. Biol. Chem. 269:12874-12879(1994).
RN   [12]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX   MEDLINE; 95172072.
RA   Schreuder H.A., Rondeau J.-M., Tardif C., Soffientini A., Sarubbi E.,
RA   Akeson A., Bowlin T.L., Yanofsky S., Barrett R.W.;
RT   "Refined crystal structure of the interleukin-1 receptor antagonist.
RT   Presence of a disulfide link and a cis-proline.";
RL   Eur. J. Biochem. 227:838-847(1995).
RN   [13]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 32-177 IN COMPLEX WITH IL1R.
RX   MEDLINE; 97215904.
RA   Schreuder H., Tardif C., Trump-Kallmeyer S., Soffientini A.,
RA   Sarubbi E., Akeson A., Bowlin T., Yanofsky S., Barrett R.W.;
RT   "A new cytokine-receptor binding mode revealed by the crystal
RT   structure of the IL-1 receptor with an antagonist.";
RL   Nature 386:194-200(1997).
CC   -!- FUNCTION: IL-1RA INHIBITS THE ACTIVITY OF IL-1 BY BINDING TO ITS
CC       RECEPTOR. IL-1RA HAS NO IL-1 LIKE ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: SECRETED OR INTRACELLULAR (THE VARIANT
CC       FORM).
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE
CC       SPLICING.
CC   -!- TISSUE SPECIFICITY: THE INTRACELLULAR FORM OF IL-1RA IS
CC       PREDOMINANTLY EXPRESSED IN EPITHELIAL CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IL-1 FAMILY.
CC   -!- DATABASE: NAME=R&D Systems' cytokine source book;
CC       WWW="HTTP://WWW.RNDSYSTEMS.COM/CYT_CAT/IL1RA.HTML".
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M55646; AAA59138.1; -.
DR   EMBL; M63099; AAB41943.1; -.
DR   EMBL; X52015; CAA36262.1; -.
DR   EMBL; X53296; CAA37386.1; -.
DR   EMBL; X64532; CAA45832.1; -.
DR   EMBL; U65590; AAB92268.1; -.
DR   EMBL; U65590; AAB92270.1; -.
DR   PIR; A30368; A30368.
DR   PIR; A37822; A37822.
DR   PIR; S08160; S08160.
DR   PIR; S08159; S08159.
DR   PIR; A40956; A40956.
DR   PIR; A39386; A39386.
DR   PDB; 1ITN; 30-APR-94.
DR   PDB; 2IRT; 15-OCT-94.
DR   PDB; 1IRP; 27-FEB-95.
DR   PDB; 1ILR; 07-FEB-95.
DR   PDB; 1ILT; 01-APR-95.
DR   PDB; 1IRA; 17-JUN-98.
DR   AARHUS/GHENT-2DPAGE; 7104; IEF.
DR   AARHUS/GHENT-2DPAGE; 7105; IEF.
DR   MIM; 147679; -.
DR   INTERPRO; IPR000975; -.
DR   PFAM; PF00340; interleukin-1; 1.
DR   PRINTS; PR00264; INTERLEUKIN1.
DR   PROSITE; PS00253; INTERLEUKIN_1; 1.
KW   Glycoprotein; Signal; Alternative splicing; 3D-structure.
FT   SIGNAL        1     25
FT   CHAIN        26    177       INTERLEUKIN-1 RECEPTOR ANTAGONIST
FT                                PROTEIN.
FT   DISULFID     91    141
FT   CARBOHYD    109    109       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC      1     21       MEICRGLRSHLITLLLFLFHS -> MAL (IN
FT                                INTRACELLULAR ISOFORM).
SQ   SEQUENCE   177 AA;  20055 MW;  D1690776A7394057 CRC64;


  Query Match             73.3%;  Score 33;  DB 1;  Length 177;
  Best Local Similarity   66.7%;  Pred. No. 9.6;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSGANLNL 9
            || | |:||
Db       59 YLQGPNVNL 67


RESULT   9
GCSB_FLAPR
ID   GCSB_FLAPR     STANDARD;      PRT;  1034 AA.
AC   P49362;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLYCINE DEHYDROGENASE [DECARBOXYLATING] B, MITOCHONDRIAL PRECURSOR
DE   (EC 1.4.4.2) (GLYCINE DECARBOXYLASE B) (GLYCINE CLEAVAGE SYSTEM P-
```

```
DE   PROTEIN B).
GN   GDCSPB.
OS   Flaveria pringlei.
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Asteridae;
OC   euasterids II; Asterales; Asteraceae; Asteroideae; Helenieae;
OC   Flaveria.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LEAF;
RX   MEDLINE; 96096729.
RA   Bauwe H., Chu C.-C., Kopriva S., Nan Q.;
RT   "Structure and expression analysis of the gdcsPA and gdcsPB genes
RT   encoding two P-isoproteins of the glycine-cleavage system from
RT   Flaveria pringlei.";
RL   Eur. J. Biochem. 234:116-124(1995).
CC   -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC       GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE
CC       THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND
CC       THE REMAINING METHYLAMINE MOIETY IS THE THEN TRANSFERRED TO THE
CC       LIPOAMIDE COFACTOR OF THE H PROTEIN.
CC   -!- CATALYTIC ACTIVITY: GLYCINE + LIPOYLPROTEIN = S-AMINOMETHYL-
CC       DIHYDROLIPOYLPROTEIN + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY). THE GLYCINE CLEAVAGE SYSTEM IS
CC       COMPOSED OF FOUR PROTEINS: P, T, L, AND H.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC   -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z54239; CAA91000.1; -.
KW   Oxidoreductase; Mitochondrion; Pyridoxal phosphate; Transit peptide;
KW   Multigene family.
FT   TRANSIT       1     63       MITOCHONDRION (POTENTIAL).
FT   CHAIN        64   1034       GLYCINE DEHYDROGENASE [DECARBOXYLATING]
FT                                B.
FT   BINDING     770    770       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   1034 AA;  112780 MW;  713D6490B48C2932 CRC64;


  Query Match             73.3%;  Score 33;  DB 1;  Length 1034;
  Best Local Similarity   62.5%;  Pred. No. 59;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANLN 8
          |: |||:|
Db    741 YMDGANMN 748


RESULT 10
GCSP_FLAAN
ID   GCSP_FLAAN     STANDARD;      PRT;  1034 AA.
AC   O49850;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLYCINE DEHYDROGENASE [DECARBOXYLATING], MITOCHONDRIAL PRECURSOR
DE   (EC 1.4.4.2) (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-
DE   PROTEIN).
GN   GDCSP.
OS   Flaveria anomala.
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Asteridae;
OC   euasterids II; Asterales; Asteraceae; Asteroideae; Helenieae;
OC   Flaveria.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LEAF;
RA   Nan Q., Bauwe H.;
RT   "The GDCSP gene encoding P-protein of the glycine cleavage system in
RT   the C3-C4 intermediate plant Flaveria anomala.";
RL   (In) Plant Gene Register PGR98-004.
CC   -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC       GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE
CC       THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND
CC       THE REMAINING METHYLAMINE MOIETY IS THE THEN TRANSFERRED TO THE
CC       LIPOAMIDE COFACTOR OF THE H PROTEIN.
CC   -!- CATALYTIC ACTIVITY: GLYCINE + LIPOYLPROTEIN = S-AMINOMETHYL-
CC       DIHYDROLIPOYLPROTEIN + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY). THE GLYCINE CLEAVAGE SYSTEM IS
CC       COMPOSED OF FOUR PROTEINS: P, T, L, AND H.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC   -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z99762; CAB16911.1; -.
KW   Oxidoreductase; Mitochondrion; Pyridoxal phosphate; Transit peptide.
FT   TRANSIT       1     63       MITOCHONDRION (POTENTIAL).
FT   CHAIN        64   1034       GLYCINE DEHYDROGENASE [DECARBOXYLATING].
FT   BINDING     770    770       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   1034 AA;  112695 MW;  C2F424C76EF104BA CRC64;

  Query Match             73.3%;  Score 33;  DB 1;  Length 1034;
  Best Local Similarity   62.5%;  Pred. No. 59;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANLN 8
          |: |||:|
Db    741 YMDGANMN 748


RESULT 11
GCSP_FLATR
ID   GCSP_FLATR     STANDARD;      PRT;  1034 AA.
AC   O49852;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLYCINE DEHYDROGENASE [DECARBOXYLATING], MITOCHONDRIAL PRECURSOR
DE   (EC 1.4.4.2) (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-
DE   PROTEIN).
GN   GDCSPA OR GDCSP.
OS   Flaveria trinervia.
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Asteridae;
OC   euasterids II; Asterales; Asteraceae; Asteroideae; Helenieae;
OC   Flaveria.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LEAF;
RA   Cossu R., Bauwe H.;
RT   "Two genes of the GDCSP gene family from the C4 plant Flaveria
RT   trinervia: GDCSPA encoding P-protein and GDCSPB, a pseudogene.";
RL   (In) Plant Gene Register PGR98-002.
CC   -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC       GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE
CC       THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND
CC       THE REMAINING METHYLAMINE MOIETY IS THE THEN TRANSFERRED TO THE
CC       LIPOAMIDE COFACTOR OF THE H PROTEIN.
CC   -!- CATALYTIC ACTIVITY: GLYCINE + LIPOYLPROTEIN = S-AMINOMETHYL-
```

```
CC       DIHYDROLIPOYLPROTEIN + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY). THE GLYCINE CLEAVAGE SYSTEM IS
CC       COMPOSED OF FOUR PROTEINS: P, T, L, AND H.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC   -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z99767; CAB16916.1; -.
KW   Oxidoreductase; Mitochondrion; Pyridoxal phosphate; Transit peptide.
FT   TRANSIT       1     63       MITOCHONDRION (POTENTIAL).
FT   CHAIN        64   1034       GLYCINE DEHYDROGENASE [DECARBOXYLATING].
FT   BINDING     770    770       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   1034 AA;  112735 MW;  14A71076C05A5062 CRC64;

  Query Match             73.3%;  Score 33;  DB 1;  Length 1034;
  Best Local Similarity   62.5%;  Pred. No. 59;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGANLN 8
              |: |||:|
Db        741 YMDGANMN 748


RESULT 12
GCSP_SOLTU
ID   GCSP_SOLTU     STANDARD;      PRT;  1035 AA.
AC   O49954;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   GLYCINE DEHYDROGENASE [DECARBOXYLATING], MITOCHONDRIAL PRECURSOR
DE   (EC 1.4.4.2) (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-
DE   PROTEIN).
GN   GDCSP.
OS   Solanum tuberosum (Potato).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC   Solanales; Solanaceae; Solanum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. DESIREE; TISSUE=LEAF;
RA   Bauwe H.;
RT   "cDNA encoding P-protein of the glycine cleavage system in Solanum
RT   tuberosum Cv. Desire.";
RL   (In) Plant Gene Register PGR98-005.
CC   -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC       GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE
CC       THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND
CC       THE REMAINING METHYLAMINE MOIETY IS THE THEN TRANSFERRED TO THE
CC       LIPOAMIDE COFACTOR OF THE H PROTEIN.
CC   -!- CATALYTIC ACTIVITY: GLYCINE + LIPOYLPROTEIN = S-AMINOMETHYL-
CC       DIHYDROLIPOYLPROTEIN + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY). THE GLYCINE CLEAVAGE SYSTEM IS
CC       COMPOSED OF FOUR PROTEINS: P, T, L, AND H.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC   -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z99770; CAB16918.1; -.
KW   Oxidoreductase; Mitochondrion; Pyridoxal phosphate; Transit peptide.
FT   TRANSIT       1     64       MITOCHONDRION (POTENTIAL).
FT   CHAIN        65   1035       GLYCINE DEHYDROGENASE [DECARBOXYLATING].
FT   BINDING     771    771       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   1035 AA;  112914 MW;  76C8418ED1856AFB CRC64;

  Query Match             73.3%;  Score 33;  DB 1;  Length 1035;
  Best Local Similarity   62.5%;  Pred. No. 59;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGANLN 8
              |: |||:|
Db        742 YMDGANMN 749


RESULT 13
GCSA_FLAPR
ID   GCSA_FLAPR     STANDARD;      PRT;  1037 AA.
AC   P49361;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLYCINE DEHYDROGENASE [DECARBOXYLATING] A, MITOCHONDRIAL PRECURSOR
DE   (EC 1.4.4.2) (GLYCINE DECARBOXYLASE A) (GLYCINE CLEAVAGE SYSTEM P-
DE   PROTEIN A).
GN   GDCSPA.
OS   Flaveria pringlei.
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Asteridae;
OC   euasterids II; Asterales; Asteraceae; Asteroideae; Helenieae;
OC   Flaveria.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LEAF;
RX   MEDLINE; 95241630.
RA   Bauwe H., Kopriva S.;
RT   "The gdcsPA gene from Flaveria pringlei (Asteraceae).";
RL   Plant Physiol. 107:655-655(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LEAF;
RX   MEDLINE; 94218395.
RA   Kopriva S., Bauwe H.;
RT   "P-protein of glycine decarboxylase from Flaveria pringlei.";
RL   Plant Physiol. 104:1077-1078(1994).
CC   -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC       GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE
CC       THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND
CC       THE REMAINING METHYLAMINE MOIETY IS THE THEN TRANSFERRED TO THE
CC       LIPOAMIDE COFACTOR OF THE H PROTEIN.
CC   -!- CATALYTIC ACTIVITY: GLYCINE + LIPOYLPROTEIN = S-AMINOMETHYL-
CC       DIHYDROLIPOYLPROTEIN + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY). THE GLYCINE CLEAVAGE SYSTEM IS
CC       COMPOSED OF FOUR PROTEINS: P, T, L, AND H.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC   -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z36879; CAA85353.1; -.
DR   EMBL; Z25857; CAA81076.1; -.
KW   Oxidoreductase; Mitochondrion; Pyridoxal phosphate; Transit peptide;
```

```
KW   Multigene family.
FT   TRANSIT       1     66       MITOCHONDRION (POTENTIAL).
FT   CHAIN        67   1037       GLYCINE DEHYDROGENASE [DECARBOXYLATING]
FT                                A.
FT   BINDING     773    773       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   CONFLICT      2      2       E -> D (IN REF. 2).
FT   CONFLICT    495    495       T -> I (IN REF. 2).
SQ   SEQUENCE   1037 AA;  113031 MW;  ED248FA227F9E0F3 CRC64;


  Query Match             73.3%;  Score 33;  DB 1;  Length 1037;
  Best Local Similarity   62.5%;  Pred. No. 59;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANLN 8
           |: |||:|
Db     744 YMDGANMN 751


RESULT  14
GCSP_PEA
ID   GCSP_PEA       STANDARD;      PRT;  1057 AA.
AC   P26969;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLYCINE DEHYDROGENASE [DECARBOXYLATING], MITOCHONDRIAL PRECURSOR
DE   (EC 1.4.4.2) (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGE SYSTEM P-
DE   PROTEIN).
GN   GDCSP OR GDCP.
OS   Pisum sativum (Garden pea).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Pisum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. BIRTE; TISSUE=LEAF;
RX   MEDLINE; 92184787.
RA   Turner S.R., Irland R., Rawsthorne S.;
RT   "Cloning and characterization of the P subunit of glycine
RT   decarboxylase from pea (Pisum sativum).";
RL   J. Biol. Chem. 267:5355-5360(1992).
RN   [2]
RP   SEQUENCE OF 905-1057 FROM N.A.
RC   STRAIN=CV. ALASKA;
RA   Shah K.S., Kim Y., Oliver D.J.;
RL   Submitted (AUG-1990) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC       GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE
CC       THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND
CC       THE REMAINING METHYLAMINE MOIETY IS THE THEN TRANSFERRED TO THE
CC       LIPOAMIDE COFACTOR OF THE H PROTEIN.
CC   -!- CATALYTIC ACTIVITY: GLYCINE + LIPOYLPROTEIN = S-AMINOMETHYL-
CC       DIHYDROLIPOYLPROTEIN + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- SUBUNIT: HOMODIMER. THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF
CC       FOUR PROTEINS: P, T, L, AND H.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC   -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X59773; CAA42443.1; -.
DR   EMBL; X54377; CAA38252.1; -.
DR   PIR; S16391; S16391.
DR   PIR; A42109; A42109.
KW   Oxidoreductase; Mitochondrion; Pyridoxal phosphate; Transit peptide.
FT   TRANSIT       1     86       MITOCHONDRION (POTENTIAL).
FT   CHAIN        87   1057       GLYCINE DEHYDROGENASE [DECARBOXYLATING].
FT   BINDING     792    792       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   CONFLICT    906    906       I -> Y (IN REF. 2).
FT   CONFLICT    919    919       P -> A (IN REF. 2).
SQ   SEQUENCE   1057 AA;  114686 MW;  2F2EA58E9A2AC447 CRC64;

  Query Match             73.3%;  Score 33;  DB 1;  Length 1057;
  Best Local Similarity   62.5%;  Pred. No. 60;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANLN 8
           |: |||:|
Db     763 YMDGANMN 770


RESULT  15
YJ89_PYRHO
ID   YJ89_PYRHO     STANDARD;      PRT;   335 AA.
AC   O57713;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PROTEIN PH1989.
GN   PH1989.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE; 98344137.
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE UPF0104 FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AP000007; BAA31116.1; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM      4     24       POTENTIAL.
FT   TRANSMEM     34     54       POTENTIAL.
FT   TRANSMEM     62     82       POTENTIAL.
FT   TRANSMEM    122    142       POTENTIAL.
FT   TRANSMEM    148    168       POTENTIAL.
FT   TRANSMEM    231    251       POTENTIAL.
FT   TRANSMEM    266    286       POTENTIAL.
FT   TRANSMEM    304    324       POTENTIAL.
SQ   SEQUENCE   335 AA;  37502 MW;  95745BC337B7B494 CRC64;


  Query Match             71.1%;  Score 32;  DB 1;  Length 335;
  Best Local Similarity   55.6%;  Pred. No. 30;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANLNL 9
           :| |||:|:
Db      59 FLKGANINV 67
```

```
Search completed: December 16, 2000, 03:05:18
Job time: 8038 sec
```

```
                    GenCore version 4.5
           Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 15, 2000, 23:43:59 ; Search time 91.14 Seconds
                                               (without alignments)
                                               6.267 Million cell updates/sec

Title:           US-09-529-121-1
Perfect score:   45
Sequence:        1 YLSGANLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters:      182106

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_65:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                   %
Result             Query
  No.    Score     Match Length DB  ID                        Description
-------------------------------------------------------------------------------
    1       45     100.0    702  2  A36319                    carcinoembryonic a
    2       39      86.7    506  2  C81704                    monooxygenase-rela
    3       35      77.8    949  2  E75352                    glycine cleavage s
    4       35      77.8    950  2  D81821                    glycine dehydrogen
    5       35      77.8    957  2  S36834                    glycine dehydrogen
    6       35      77.8    979  2  T16734                    hypothetical prote
    7       35      77.8   1003  2  A39521                    glycine dehydrogen
    8       35      77.8   1020  2  JN0124                    glycine dehydrogen
    9       35      77.8   1020  2  B39521                    glycine dehydrogen
   10       35      77.8   1034  2  S50917                    aminomethyltransfe
   11       34      75.6    259  2  S76576                    hypothetical prote
   12       34      75.6    272  2  T22562                    hypothetical prote
   13       34      75.6    473  2  B81919                    probable membrane
   14       34      75.6    473  2  E81182                    hypothetical prote
   15       34      75.6    554  2  A70904                    probable acid--CoA
   16       34      75.6    941  2  A70722                    probable gcvB prot
   17       34      75.6    952  2  T44754                    probable glycine d
   18       33      73.3     85  2  S08109                    carcinoembryonic a
   19       33      73.3    177  2  A30368                    interleukin-1 rece
   20       33      73.3    180  2  A39386                    interleukin-1 rece
   21       33      73.3    356  2  T28747                    hypothetical prote
   22       33      73.3    507  2  E71551                    probable monooxyge
   23       33      73.3    983  2  S76257                    hypothetical prote
   24       33      73.3   1031  2  T46636                    glycine dehydrogen
   25       33      73.3   1034  2  S63536                    aminomethyltransfe
   26       33      73.3   1035  2  T07826                    aminomethyltransfe
   27       33      73.3   1037  2  T05309                    probable glycine d
   28       33      73.3   1037  2  S63535                    aminomethyltransfe
   29       33      73.3   1037  2  S40216                    P protein - Flaver
```

```
   30       33      73.3   1044  2  T02615                    probable glycine d
   31       33      73.3   1057  2  A42109                    glycine dehydrogen
   32       33      73.3   1180  2  S69205                    stripe a/b protein
   33       33      73.3   1519  2  S41525                    major ring-forming
   34       32      71.1     83  2  I64001                    hypothetical prote
   35       32      71.1    236  2  T22220                    hypothetical prote
   36       32      71.1    335  2  E71215                    hypothetical prote
   37       32      71.1    355  2  G64404                    hypothetical prote
   38       32      71.1    388  2  S74425                    hypothetical prote
   39       32      71.1    455  2  S48859                    nucleoside triphos
   40       32      71.1    474  2  H72403                    glycine dehydrogen
   41       32      71.1    488  2  B69959                    glycine dehydrogen
   42       32      71.1    502  2  B71216                    probable glycine d
   43       32      71.1    502  2  D75030                    probable glycine d
   44       32      71.1    521  2  D72518                    probable glycine d
   45       32      71.1    576  1  DWBYT                     threonine dehydrat

                                  ALIGNMENTS


RESULT   1
A36319
carcinoembryonic antigen precursor - human
N;Alternate names: CEA; meconium antigen 100
C;Species: Homo sapiens (man)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 31-Jan-2000
C;Accession: A36319; A27773; A31037; A25845; S08106; S31737; A44476; I54224; I59098;
R;Schrewe, H.; Thompson, J.; Bona, M.; Hefta, L.J.F.; Maruya, A.; Hassauer, M.; Shive
Mol. Cell. Biol. 10, 2738-2748, 1990
A;Title: Cloning of the complete gene for carcinoembryonic antigen: analysis of its p
A;Reference number: A36319; MUID:90258861
A;Accession: A36319
A;Molecule type: DNA
A;Residues: 1-702 <SCH>
A;Cross-references: GB:M17303; NID:g178676; PIDN:AAB59513.1; PID:g178677
A;Note: the authors show the codons TTA for residue 641-Phe and CAG for residue 646-T
R;Beauchemin, N.; Benchimol, S.; Cournoyer, D.; Fuks, A.; Stanners, C.P.
Mol. Cell. Biol. 7, 3221-3230, 1987
A;Title: Isolation and characterization of full-length functional cDNA clones for hum
A;Reference number: A27773; MUID:88038876
A;Accession: A27773
A;Molecule type: mRNA
A;Residues: 1-702 <BEA>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
R;Barnett, T.; Goebel, S.J.; Nothdurft, M.A.; Elting, J.J.
Genomics 3, 59-66, 1988
A;Title: Carcinoembryonic antigen family: characterization of cDNAs coding for NCA an
A;Reference number: A31037; MUID:89122014
A;Accession: A31037
A;Molecule type: mRNA
A;Residues: 1-702 <BAR>
A;Cross-references: GB:M29540; NID:g180222; PIDN:AAA51967.1; PID:g180223
A;Note: the authors translated the codon GTG for residue 130 as Leu
R;Oikawa, S.; Nakazato, H.; Kosaki, G.
Biochem. Biophys. Res. Commun. 142, 511-518, 1987
A;Title: Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA
A;Reference number: A25845; MUID:87128144
A;Accession: A25845
A;Molecule type: mRNA
A;Residues: 5-702 <OIK>
A;Cross-references: GB:M15042; NID:g180198; PIDN:AAA51963.1; PID:g180199
R;Oikawa, S.
submitted to the EMBL Data Library, September 1989
A;Reference number: S08106
A;Accession: S08106
A;Molecule type: mRNA
A;Residues: 5-319,321-702 <OI2>
A;Cross-references: EMBL:X16455; NID:g29854; PIDN:CAA34474.1; PID:g825638
R;Barnett, T.
submitted to the EMBL Data Library, September 1991
A;Description: Genomic DNA sequence upstream of the translational start of the carcin
A;Reference number: S31737
```

```
A;Accession: S31737
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-141 <BA2>
A;Cross-references: EMBL:X62151
R;Khan, W.N.; Fraengsmyr, L.; Teglund, S.; Israelsson, A.; Bremer, K.; Hammarstroem, S.
Genomics 14, 384-390, 1992
A;Title: Identification of three new genes and estimation of the size of the carcinoembr
A;Reference number: A44476; MUID:93052339
A;Accession: A44476
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 35-141 <KHA>
R;Willcocks, T.C.; Craig, I.W.
Genomics 8, 492-500, 1990
A;Title: Characterization of the genomic organization of human carcinoembryonic antigen
A;Reference number: I54224; MUID:91139118
A;Accession: I54224
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-37 <RES>
A;Cross-references: GB:M60964; NID:g180215; PIDN:AAA51964.1; PID:g180217
R;Zimmermann, W.; Ortlieb, B.; Friedrich, R.; von Kleist, S.
Proc. Natl. Acad. Sci. U.S.A. 84, 2960-2964, 1987
A;Title: Isolation and characterization of cDNA clones encoding the human carcinoembryon
A;Reference number: I59098; MUID:87204247
A;Accession: I59098
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 331-702 <RE2>
A;Cross-references: GB:M16234; NID:g180240; PIDN:AAA51972.1; PID:g180241
R;Siepen, D.; Paxton, R.J.; Neumaier, M.; Shively, J.E.; Wagener, C.
Biochem. Biophys. Res. Commun. 147, 212-218, 1987
A;Title: Carcinoembryonic antigen (CEA) and two crossreacting antigens of 165 KD and 105
A;Reference number: A26831; MUID:87326349
A;Accession: A26831
A;Molecule type: protein
A;Residues: 35-64 <SIE>
R;Thomas, P.; Toth, C.A.
Biochem. Biophys. Res. Commun. 170, 391-396, 1990
A;Title: Carcinoembryonic antigen binding to Kupffer cells is via a peptide located at t
A;Reference number: A35490; MUID:90321257
A;Accession: A35490
A;Molecule type: protein
A;Residues: 'X',140-151,'X',153,'X',155-156 <THO>
A;Note: this is the amino terminal end of a fragment shown to mediate uptake by Kupffer
C;Comment: This heavily glycosylated membane protein of unknown function is a widely use
C;Comment: This protein may be processed at its C-terminus. It is anchored to the membra
C;Genetics:
A;Gene: GDB:CEA
A;Cross-references: GDB:119054; OMIM:114890
A;Map position: 19q13.2-19q13.2
A;Introns: 22/1; 142/1; 235/1; 320/1; 413/1; 498/1; 591/1; 676/1
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin
C;Keywords: blocked carboxyl end; glycoprotein; lipoprotein; membrane protein; phosphati
F;1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-678/Product: carcinoembryonic antigen #status predicted <MAT>
F;160-217/Domain: immunoglobulin homology <IMM1>
F;252-301/Domain: immunoglobulin homology <IMM2>
F;338-395/Domain: immunoglobulin homology <IMM3>
F;516-573/Domain: immunoglobulin homology <IMM4>
F;608-657/Domain: immunoglobulin homology <IMM5>
F;679-702/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F;678/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Gly) (in mature form


  Query Match             100.0%;  Score 45;  DB 2;  Length 702;
  Best Local Similarity   100.0%;  Pred. No. 0.21;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YLSGANLNL 9
                |||||||||
Db        605 YLSGANLNL 613


RESULT   2
C81704
monooxygenase-related protein TC0425 [imported] - Chlamydia muridarum (strain Nigg)
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C;Accession: C81704
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A;Reference number: A81500; MUID:20150255
A;Accession: C81704
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-506 <TET>
A;Cross-references: GB:AE002309; GB:AE002160; NID:g7190464; PIDN:AAF39281.1; PID:g719
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
A;Gene: TC0425


  Query Match             86.7%;  Score 39;  DB 2;  Length 506;
  Best Local Similarity   77.8%;  Pred. No. 2.7;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGANLNL 9
                |||| |||:
Db        290 YLSGVNLNI 298


RESULT   3
E75352
glycine cleavage system P protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: E75352
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: E75352
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-949 <WHI>
A;Cross-references: GB:AE002021; GB:AE000513; NID:g6459573; PIDN:AAF11360.1; PID:g645
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR1809
A;Map position: 1


  Query Match             77.8%;  Score 35;  DB 2;  Length 949;
  Best Local Similarity   75.0%;  Pred. No. 36;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGANLN 8
                || |||:|
Db        668 YLDGANMN 675


RESULT   4
D81821
glycine dehydrogenase (EC 1.4.4.2) NMA1934 [imported] - Neisseria meningitidis (group
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
C;Accession: D81821
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
```

```
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491.
A;Reference number: A81775; MUID:20222556
A;Accession: D81821
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-950 <PAR>
A;Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB85154.1; PID:g738056
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: gcvP; NMA1934
C;Keywords: oxidoreductase


  Query Match             77.8%;  Score 35;  DB 2;  Length 950;
  Best Local Similarity   75.0%;  Pred. No. 36;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANLN 8
             |: |||||
Db     669 YMDGANLN 676


RESULT   5
S36834
glycine dehydrogenase (decarboxylating) (EC 1.4.4.2) - Escherichia coli
C;Species: Escherichia coli
C;Date: 22-Jan-1994 #sequence_revision 10-Nov-1995 #text_change 08-Oct-1999
C;Accession: S36834; I41232; G65074
R;Okamura-Ikeda, K.; Ohmura, Y.; Fujiwara, K.; Motokawa, Y.
Eur. J. Biochem. 216, 539-548, 1993
A;Title: Cloning and nucleotide sequence of the gcv operon encoding the Escherichia coli
A;Reference number: S36832; MUID:93387305
A;Accession: S36834
A;Molecule type: DNA
A;Residues: 1-957 <OKA>
A;Cross-references: EMBL:X73958; NID:g403342; PIDN:CAA52146.1; PID:g403345
R;Stauffer, L.T.; Fogarty, S.J.; Stauffer, G.V.
Gene 142, 17-22, 1994
A;Title: Characterization of the Escherichia coli gcv operon.
A;Reference number: I41231; MUID:94237484
A;Accession: I41232
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-957 <RES>
A;Cross-references: GB:L20872; NID:g304890; PIDN:AAA23867.1; PID:g304892
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: G65074
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-957 <BLAT>
A;Cross-references: GB:AE000373; GB:U00096; NID:g2367173; PIDN:AAC75941.1; PID:g1789269;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: gcvP; gcvHP
C;Keywords: oxidoreductase; phosphoprotein; pyridoxal phosphate
F;708/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted


  Query Match             77.8%;  Score 35;  DB 2;  Length 957;
  Best Local Similarity   75.0%;  Pred. No. 36;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANLN 8
             || |||:|
Db     679 YLDGANMN 686


RESULT   6
T16734
hypothetical protein R12C12.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T16734
R;Favello, T.
submitted to the EMBL Data Library, July 1995
A;Description: The sequence of C. elegans cosmid R12C12.
A;Reference number: Z18568
A;Accession: T16734
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-979 <FAV>
A;Cross-references: EMBL:U23510; NID:g746453; PID:g746454; PIDN:AAC46780.1; CESP:R12C
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:R12C12.1
A;Introns: 52/3; 149/3; 325/2; 393/1; 470/3; 546/3; 619/3; 749/2; 827/1; 883/1


  Query Match             77.8%;  Score 35;  DB 2;  Length 979;
  Best Local Similarity   75.0%;  Pred. No. 37;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANLN 8
             || |||:|
Db     702 YLDGANMN 709


RESULT   7
A39521
glycine dehydrogenase (decarboxylating) (EC 1.4.4.2) precursor - chicken
N;Alternate names: glycine decarboxylase; P-protein
C;Species: Gallus gallus (chicken)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 05-Dec-1998
C;Accession: A39521; C39521; A27483
R;Kume, A.; Koyata, H.; Sakakibara, T.; Ishiguro, Y.; Kure, S.; Hiraga, K.
J. Biol. Chem. 266, 3323-3329, 1991
A;Title: The glycine cleavage system. Molecular cloning of the chicken and human glyc
A;Reference number: A39521; MUID:91131643
A;Accession: A39521
A;Molecule type: mRNA
A;Residues: 1-1003 <KUM>
A;Cross-references: GB:D90266
A;Accession: C39521
A;Molecule type: DNA
A;Residues: 1-13 <KU2>
A;Cross-references: GB:D90240
A;Note: part of this sequence was confirmed by protein sequencing
R;Fujiwara, K.; Okamura-Ikeda, K.; Motokawa, Y.
Biochem. Biophys. Res. Commun. 149, 621-627, 1987
A;Title: Amino acid sequence of the phosphopyridoxyl peptide from P-protein of the ch
A;Reference number: A27483; MUID:88106483
A;Accession: A27483
A;Molecule type: protein
A;Residues: 703-756 <FUJ>
C;Keywords: mitochondrion; oxidoreductase; phosphoprotein; pyridoxal phosphate
F;737/Binding site: pyridoxal phosphate (Lys) (covalent) #status experimental


  Query Match             77.8%;  Score 35;  DB 2;  Length 1003;
  Best Local Similarity   75.0%;  Pred. No. 38;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANLN 8
             || |||:|
Db     708 YLDGANMN 715


RESULT   8
JN0124
glycine dehydrogenase (decarboxylating) (EC 1.4.4.2) - human
N;Alternate names: glycine decarboxylase; P-protein
```

```
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-Nov-1999
C;Accession: JN0124
R;Kure, S.; Narisawa, K.; Tada, K.
Biochem. Biophys. Res. Commun. 174, 1176-1182, 1991
A;Title: Structural and expression analyses of normal and mutant mRNA encoding glycine d
A;Reference number: JN0124; MUID:91144593
A;Accession: JN0124
A;Molecule type: mRNA
A;Residues: 1-1020 <KUR>
A;Cross-references: GB:M63635; NID:g190286; PIDN:AAA36478.1; PID:g190287
A;Note: deletion of 756-Phe causes nonketotic hyperglycinemia
C;Genetics:
A;Gene: GDB:GLDC
A;Cross-references: GDB:128611; OMIM:238300
A;Map position: 9p22-9p22
C;Keywords: mitochondrion; oxidoreductase; phosphoprotein; pyridoxal phosphate
F;761-768/Region: glycine-rich
F;754/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

  Query Match             77.8%;  Score 35;  DB 2;  Length 1020;
  Best Local Similarity   75.0%;  Pred. No. 39;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANLN 8
          || |||:|
Db    725 YLDGANMN 732


RESULT   9
B39521
glycine dehydrogenase (decarboxylating) (EC 1.4.4.2) precursor - human
N;Alternate names: glycine decarboxylase; P-protein
C;Species: Homo sapiens (man)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 05-Nov-1999
C;Accession: B39521
R;Kume, A.; Koyata, H.; Sakakibara, T.; Ishiguro, Y.; Kure, S.; Hiraga, K.
J. Biol. Chem. 266, 3323-3329, 1991
A;Title: The glycine cleavage system. Molecular cloning of the chicken and human glycine
A;Reference number: A39521; MUID:91131643
A;Accession: B39521
A;Molecule type: mRNA
A;Residues: 1-1020 <KUM>
A;Cross-references: GB:M64590; GB:J05742; NID:g190208; PIDN:AAA36463.1; PID:g190209
C;Genetics:
A;Gene: GDB:GLDC
A;Cross-references: GDB:128611; OMIM:238300
A;Map position: 9p22-9p22
C;Keywords: mitochondrion; oxidoreductase; phosphoprotein; pyridoxal phosphate
F;754/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

  Query Match             77.8%;  Score 35;  DB 2;  Length 1020;
  Best Local Similarity   75.0%;  Pred. No. 39;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANLN 8
          || |||:|
Db    725 YLDGANMN 732


RESULT  10
S50917
aminomethyltransferase (EC 2.1.2.10) - yeast (Saccharomyces cerevisiae)
N;Alternate names: glycin cleavage system protein P; protein YM9646.01; protein YMR189c
C;Species: Saccharomyces cerevisiae
C;Date: 10-Feb-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C;Accession: S50917; S59810; S70896
R;Pearson, D.; Bowman, S.
submitted to the EMBL Data Library, January 1995
A;Reference number: S50917
A;Accession: S50917
```

```
A;Molecule type: DNA
A;Residues: 1-1034 <PEA>
A;Cross-references: EMBL:Z47815; NID:g642280; PIDN:CAA87810.1; PID:g642281; MIPS:YMR1
R;Sinclair, D.A.; Dawes, I.W.
submitted to the EMBL Data Library, February 1995
A;Description: Cloning and expression of glycine decarboxylase from Saccharomyces cer
A;Reference number: S59809
A;Accession: S59810
A;Molecule type: DNA
A;Residues: 1-1034 <SIN>
A;Cross-references: EMBL:U20641; NID:g676869; PIDN:AAB18933.1; PID:g676871
R;Sinclair, D.A.; Hong, S.P.; Dawes, I.W.
Mol. Microbiol. 19, 611-623, 1996
A;Title: Specific induction by glycine of the gene for the P-subunit of glycine decar
A;Reference number: S70896; MUID:96228709
A;Accession: S70896
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-77,'V',79-121,'VS',124-247,'L',249-381,'T',383-517,'H',519-1034 <SIW>
A;Cross-references: EMBL:U20641
C;Genetics:
A;Gene: GCV2; GSD2
A;Map position: 13R
C;Keywords: phosphoprotein; pyridoxal phosphate; transferase
F;773/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

  Query Match             77.8%;  Score 35;  DB 2;  Length 1034;
  Best Local Similarity   75.0%;  Pred. No. 39;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 YLSGANLN 8
          || |||:|
Db    744 YLDGANMN 751


RESULT  11
S76576
hypothetical protein sll0183 - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S76576
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S76576
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-259 <KAN>
A;Cross-references: EMBL:D64002; GB:AB001339; NID:g1001612; PIDN:BAA10422.1; PID:d101
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

  Query Match             75.6%;  Score 34;  DB 2;  Length 259;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 LSGANLN 8
          |||||||
Db     91 LSGANLN 97


RESULT  12
T22562
hypothetical protein F53F1.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T22562
R;Burton, J.
```

```
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19583
A;Accession: T22562
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-272 <WIL>
A;Cross-references: EMBL:Z81088; PIDN:CAB03123.1; GSPDB:GN00023; CESP:F53F1.6
A;Experimental source: clone F53F1
C;Genetics:
A;Gene: CESP:F53F1.6
A;Map position: 5
A;Introns: 19/3; 54/3; 122/2; 149/3; 161/3; 236/3


  Query Match             75.6%;  Score 34;  DB 2;  Length 272;
  Best Local Similarity   66.7%;  Pred. No. 15;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSGANLNL 9
             || | |||:
Db     235 YLEGVNLNV 243


RESULT  13
B81919
probable membrane protein NMA0753 [imported] - Neisseria meningitidis (group A strain Z2
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
C;Accession: B81919
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491.
A;Reference number: A81775; MUID:20222556
A;Accession: B81919
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-473 <PAR>
A;Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84036.1; PID:g737947
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: NMA0753


  Query Match             75.6%;  Score 34;  DB 2;  Length 473;
  Best Local Similarity   87.5%;  Pred. No. 27;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LSGANLNL 9
             | ||||||
Db      21 LDGANLNL 28


RESULT  14
E81182
hypothetical protein NMB0570 [imported] - Neisseria meningitidis (group B strain MD58)
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C;Accession: E81182
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.;
Science 287, 1809-1815, 2000                                                            ;
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755
A;Accession: E81182
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-473 <TET>
A;Cross-references: GB:AE002413; GB:AE002098; NID:g7225797; PIDN:AAF40998.1; PID:g722579
A;Experimental source: serogroup B, strain MMD58
C;Genetics:
A;Gene: NMB0570

  Query Match             75.6%;  Score 34;  DB 2;  Length 473;
  Best Local Similarity   87.5%;  Pred. No. 27;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LSGANLNL 9
             | ||||||
Db      21 LDGANLNL 28


RESULT  15
A70904
probable acid--CoA ligase (EC 6.2.1.-) - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70904
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: A70904
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-554 <COL>
A;Cross-references: GB:Z97050; GB:AL123456; NID:g3256008; PIDN:CAB09749.1; PID:g22135
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: fadD5
C;Superfamily: 4-coumarate--CoA ligase; acetate--CoA ligase homology
C;Keywords: acid-thiol ligase
F;72-525/Domain: acetate--CoA ligase homology <ACL>


  Query Match             75.6%;  Score 34;  DB 2;  Length 554;
  Best Local Similarity   75.0%;  Pred. No. 32;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANLN 8
             | ||||:|
Db     217 YTSGANIN 224


Search completed: December 16, 2000, 01:51:13
Job time: 7634 sec
```

```
                    GenCore version 4.5
             Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 15, 2000, 19:38:20 ; Search time 108.84 Seconds
                                              (without alignments)
                                              2.827 Million cell updates/sec

Title:           US-09-529-121-1
Perfect score:   45
Sequence:        1 YLSGANLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters:      268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_36:*
                1:  /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
                2:  /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
                3:  /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
                4:  /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
                5:  /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
                6:  /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
                7:  /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
                8:  /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
                9:  /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
                10:  /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
                11:  /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
                12:  /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
                13:  /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
                14:  /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
                15:  /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
                16:  /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
                17:  /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
                18:  /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
                19:  /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
                20:  /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
                21:  /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                   %
Result           Query
  No.   Score    Match Length DB  ID                       Description
 ---------------------------------------------------------------------
    1      45    100.0      9  18  W39723                   Human carcina-embr
    2      45    100.0      9  19  W77134                   CEA synthetic pept
    3      45    100.0      9  19  W70045                   CEA derived HLA-A2
    4      45    100.0      9  20  Y47655                   Immunogenic peptid
    5      45    100.0      9  20  Y09525                   Carcinoembryonic a
    6      45    100.0     10  20  Y46555                   Immunogenic peptid
    7      45    100.0    107  20  W86133                   Protein sequence o
    8      45    100.0    178  10  P93499                 , Sequence of carcin
    9      45    100.0    468  16  R77436                   BGP (1-314)/CEA (4
   10      45    100.0    493  16  R77435                   BGP (1-314)/CEA (4
   11      45    100.0    509  16  R77437                   BGP (1-314)/CEA (4
   12      45    100.0    511  16  R77438                   BGP (1-314)/CEA (4
   13      45    100.0    642  15  R60619                   Carcinoembryonic a
   14      45    100.0    663  17  R98519                   Immunogenic carcin
   15      45    100.0    698   9  P81229                  Carcinoembryonic a
   16      45    100.0    698  16  R65168                   Carcinoembryonic a
   17      45    100.0    698  18  W22844                   Human carcinoembry
   18      45    100.0    702   9  P81222                  Carcinoembryonic a
   19      45    100.0    702  10  P94014                   Carcinoembryonic c
   20      45    100.0    702  10  P93999                   Amino acid sequenc
   21      45    100.0    702  15  R54713                   Carcinoembryonic a
   22      45    100.0    702  17  W06872                   Carcinoembryonic a
   23      45    100.0    702  20  W83137                   CEA protein.  Homo
   24      45    100.0    734  17  W00182                   Carcinoembryonic a
   25      43     95.6      9  20  Y09528                   Carcinoembryonic a
   26      42     93.3      9  21  Y54173                   HLA binding peptid
   27      40     88.9      9  20  Y09526                   Carcinoembryonic a
   28      38     84.4      9  17  W00680                   Peptide comprising
   29      38     84.4      9  20  Y09527                   Carcinoembryonic a
   30      36     80.0      9  20  Y09529                   Carcinoembryonic a
   31      35     77.8    973  18  W11853                   Hepadnavirus recep
   32      33     73.3     90  10  P93625                   Sequence of interl
   33      33     73.3     90  14  R35484                   IL-1i fragment.  H
   34      33     73.3    152  13  R27495                   Native IL-1ra poly
   35      33     73.3    152  14  R35486                   IL-1 inhibitor (IL
   36      33     73.3    153  18  W22894                   Recombinant human
   37      33     73.3    153  19  W61149                   Recombinant human
   38      33     73.3    159  12  R15262                   Variant Il-1 cytok
   39      33     73.3    159  16  R73642                   iCIL-1ra.  Homo sa
   40      33     73.3    159  19  W80777                   Human intracellula
   41      33     73.3    159  19  W37787                   Recombinant interl
   42      33     73.3    159  20  Y43532                   Cytoplasmic form o
   43      33     73.3    159  20  Y33275                   icIRAP conserved p
   44      33     73.3    159  20  Y33277                   IcIRAP conserved p
   45      33     73.3    159  20  Y28287                   Polypeptide sequen

                                 ALIGNMENTS

RESULT   1
W39723
ID   W39723 standard; peptide; 9 AA.
XX
AC   W39723;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Human carcina-embryonic antigen (CEA) peptide (pos. 571-579).
XX
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
XX
OS   Homo sapiens.
XX
PN   WO9741440-A1.
XX
PD   06-NOV-1997.
XX
PF   28-APR-1997;   97WO-NL00229.
XX
PR   23-DEC-1996;   96EP-0203670.
PR   26-APR-1996;   96EP-0201145.
XX
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
PI   Kast WM,  Melief CJM,  Offringa R,  Toes REM,  Van Der Burg SH;
XX
DR   WPI; 1997-549891/50.
XX
PT   Method of selecting T cell peptide epitope(s) - by measuring the
PT   stability of HLA class I-peptide complexes on intact B cells
XX
```

```
PS   Example 3; Page 85; 109pp; English.
XX
CC   Peptides W39430-W39734 are used in a novel method for the selection of
CC   immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC   method involves the identification of peptide sequences capable of
CC   binding to an HLA (human leukocyte antigen) class I molecule and
CC   measuring the binding of this epitope peptide to the HLA class I peptide.
CC   The stability of binding of the peptide and MHC (major histocompatablilty
CC   complex) class I molecule is measured on intact human B cells carrying
CC   the MHC molecule at their cell surfaces. The method can be used to select
CC   peptide epitopes for generating vaccines against a disease associated
CC   with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC   especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC   immune responses. Peptide W39723 is derived from the human
CC   carcino-embryonic antigen (CEA) and has the ability to bind to the human
CC   MHC Class I allele HLA-A2.1.
XX
SQ   Sequence   9 AA;


  Query Match             100.0%;  Score 45;  DB 18;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANLNL 9
           |||||||||
Db       1 ylsganlnl 9


RESULT   2
W77134
ID   W77134 standard; peptide; 9 AA.
XX
AC   W77134;
XX
DT   16-NOV-1998  (first entry)
XX
DE   CEA synthetic peptide epitope 1.
XX
KW   Tyrosinase; tyrosinase cytotoxic lymphocyte response;
KW   cytotoxic T lymphocyte; cysteine-depleted; melanoma.
XX
OS   Synthetic.
XX
PN   WO9833810-A2.
XX
PD   06-AUG-1998.
XX
PF   29-JAN-1998;   98WO-US01592.
XX
PR   30-JAN-1997;   97US-0037781.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI   Engelhard VH,  Hunt DF,  Kittlesen D,  Slingluff CL;
XX
DR   WPI; 1998-437388/37.
XX
PT   Disease specific immunogen  - comprises disease specific cytotoxic T
PT   lymphocyte epitope used to elicit melanoma specific CTL response
XX
PS   Disclosure; Page 27; 93pp; English.
XX
CC   The peptide epitope W77119-W77138 were created for human tumour-specific
CC   cytotoxic T lymphocyte response.  These peptides are are cysteine-
CC   depleted mutants of a native disease-specific CTL epitope.  The cysteine-
CC   depleted CTL epitopes elicit a stronger or more specific CTL response
CC   than the native epitope.  The epitopes can be used in a disease-specific
CC   immunogen to protect a mammal against disease in particular melanomas.
CC   The peptides may also be used to screen a sample for the presence of
CC   an antigen with the same epitope, or with a different cross-reactive
CC   epitope.
XX
```

```
SQ   Sequence   9 AA;

  Query Match             100.0%;  Score 45;  DB 19;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANLNL 9
           |||||||||
Db       1 ylsganlnl 9


RESULT   3
W70045
ID   W70045 standard; peptide; 9 AA.
XX
AC   W70045;
XX
DT   22-OCT-1998  (first entry)
XX
DE   CEA derived HLA-A2.1 binding peptide 2 (residues 605-613).
XX
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW   antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW   viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW   fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9833888-A1.
XX
PD   06-AUG-1998.
XX
PF   30-JAN-1998;   98WO-US01959.
XX
PR   31-JAN-1997;   97US-0036696.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Celis E,  Sette A,  Sidney J,  Southwood S,  Tsai V;
XX
DR   WPI; 1998-437445/37.
XX
PT   Production of antigen-specific cytotoxic T cells - by incubating
PT   immunogenic peptide(s) from antigen that binds class I major
PT   histocompatibility complex molecules with pre-treated antigen
PT   presenting cells
XX
PS   Example 6; Page 75; 104pp; English.
XX
CC   Sequences shown in W70044 to W70052 represent peptides derived from
CC   carcinoembryonic antigen (CEA). The peptides can bind to a human
CC   leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method
CC   of invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC   vitro. The method comprises contacting immunogenic peptides from an
CC   antigen that binds class I major histocompatibility complex (MHC)
CC   molecules with antigen presenting cells (APCs) pretreated with
CC   pretreatment growth factors, and incubating the APCs with purified CD8
CC   cells in the presence of at least 2 incubation growth factors, thereby
CC   producing antigen-specific CTLs. A method for specifically killing
CC   target cells in a human patient is also provided which comprises
CC   obtaining a fluid sample containing CTLs from a patient, contacting the
CC   cytotoxic T cells with APCs pretreated with pre-treatment growth factors,
CC   where the APCs comprise class I MHC molecules. The pretreated APCs are
CC   incubated with the cytotoxic growth factors, thereby producing activated
CC   CTLs which are contacted with a carrier to form a composition. The
CC   composition can then be administered to the patient. The activated CTLs
CC   can be used for treating cancers, immune disorders, viral infections,
CC   AIDS, hepatitis, bacterial infection, fungal infection, malaria or
CC   tuberculosis.
XX
```

```
SQ   Sequence   9 AA;

  Query Match             100.0%;  Score 45;  DB 19;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGANLNL 9
            |||||||||
Db        1 ylsganlnl 9


RESULT   4
Y47655
ID   Y47655 standard; Peptide; 9 AA.
XX
AC   Y47655;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2266.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 118; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;

  Query Match             100.0%;  Score 45;  DB 20;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGANLNL 9
            |||||||||
Db        1 ylsganlnl 9


RESULT   5
Y09525
ID   Y09525 standard; peptide; 9 AA.
XX
AC   Y09525;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist CAP-1.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 1; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             100.0%;  Score 45;  DB 20;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGANLNL 9
            |||||||||
Db        1 ylsganlnl 9
```

```
RESULT   6
Y46555
ID   Y46555 standard; Peptide; 10 AA.
XX
AC   Y46555;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #1166.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 76; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   10 AA;


  Query Match             100.0%;  Score 45;  DB 20;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.0043;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGANLNL 9
            |||||||||
Db        2 ylsganlnl 10


RESULT   7
W86133
ID   W86133 standard; Protein; 107 AA.
XX
AC   W86133;
XX
DT   03-MAR-1999  (first entry)
XX
DE   Protein sequence of vaccine 2 708 Vl.
XX
KW   Non-immunogenic; epitope; T-cell; immunogenicity; immune system; SK;
KW   immunoglobulin; therapeutic; streptokinase; vaccine; 708.
XX
OS   Homo sapiens.
XX
PN   WO9852976-A1.
XX
PD   26-NOV-1998.
XX
PF   21-MAY-1998;   98WO-GB01473.
XX
PR   14-APR-1998;   98GB-0007751.
PR   21-MAY-1997;   97GB-0010480.
PR   31-JUL-1997;   97GB-0016197.
PR   28-NOV-1997;   97GB-0025270.
PR   02-DEC-1997;   97US-0067235.
XX
PA   (BIOV-) BIOVATION LTD.
XX
PI   Carr FJ;
XX
DR   WPI; 1999-045301/04.
XX
PT   Reducing immunogenicity of proteins - by modifying the amino acid
PT   sequence of the protein to eliminate potential epitopes for T-cells
PT   of a given species
XX
PS   Example 4; Fig 19; 77pp; English.
XX
CC   The invention relates to a method for the production of non-immunogenic
CC   proteins. The method comprises determining at least part of the amino
CC   acid sequence of the protein; (b) identifying in the amino acid sequence
CC   one or more potential epitopes for T-cells (T-cell epitopes) of the given
CC   species; and (c) modifying the amino acid sequence to eliminate at least
CC   one of the T-cell epitopes identified in step (b) thereby to eliminate or
CC   reduce the immunogenicity of the protein when exposed to the immune
CC   system of the given species. A method of analysing a pre-existing protein
CC   to predict the basis for immunogenic responses is also provided. The
CC   methods can be used particularly for reducing the immunogenicity of
CC   immunoglobulins or therapeutic proteins, e.g. Streptokinase (SK). The
CC   products can be used for diagnosis and therapy. The present sequence
CC   represents the protein sequence of vaccine 2 708 Vl.
XX
SQ   Sequence   107 AA;


  Query Match             100.0%;  Score 45;  DB 20;  Length 107;
  Best Local Similarity   100.0%;  Pred. No. 0.06;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGANLNL 9
            |||||||||
Db       89 ylsganlnl 97


RESULT   8
P93499
ID   P93499 standard; protein; 178 AA.
XX
AC   P93499;
XX
DT   08-MAY-1990  (first entry)
XX
DE   Sequence of carcinoembryonic antigen domain III.
```

```
XX
KW   Carcinoembryonic antigen; domain III; domain A; domain B.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..89
FT                   /note="domain A"
FT   Domain          90..178
FT                   /note="domain B"
XX
PN   EP343946-A.
XX
PD   29-NOV-1989.
XX
PF   24-MAY-1989;   89EP-0305232.
XX
PR   25-MAY-1988;   88US-0198289.
XX
PA   (CITY ) CITY OF HOPE.
XX
PI   Shively JE;
XX
DR   WPI; 1989-349991/48.
DR   N-PSDB; N92449.
XX
PT   Carcinoembryonic antigen fragments - used in assays to determine the
PT   presence and amt. of the antigen in samples also contg. related antigens.
XX
PS   Disclosure; page 4; 15pp; English.
XX
CC   CEA fragments can be used in assays to determine the presence and amt. of
CC   CEA in samples which also may contain related antigens including its
CC   normal cross-reacting antigen or the 128 kD antigen.
XX
SQ   Sequence   178 AA;

  Query Match             100.0%;  Score 45;  DB 10;  Length 178;
  Best Local Similarity   100.0%;  Pred. No. 0.11;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANLNL 9
             |||||||||
Db     107 ylsganlnl 115

RESULT   9
R77436
ID   R77436 standard; Protein; 468 AA.
XX
AC   R77436;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-643) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..468
FT                   /note= "CEA (490-643)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
```

```
XX
PR   21-AUG-1993;   93GB-0017423.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 16; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   468 AA;

  Query Match             100.0%;  Score 45;  DB 16;  Length 468;
  Best Local Similarity   100.0%;  Pred. No. 0.31;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANLNL 9
             |||||||||
Db     396 ylsganlnl 404

RESULT  10
R77435
ID   R77435 standard; Protein; 493 AA.
XX
AC   R77435;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-C-terminal) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..493
FT                   /note= "CEA (490-C-terminal)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
```

```
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 15; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   493 AA;


  Query Match             100.0%;  Score 45;  DB 16;  Length 493;
  Best Local Similarity   100.0%;  Pred. No. 0.33;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGANLNL 9
            |||||||||
Db      396 ylsganlnl 404


RESULT  11
R77437
ID   R77437 standard; Protein; 509 AA.
XX
AC   R77437;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-644)/BGP (391-430) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..469
FT                   /note= "CEA (490-644)"
FT   Protein         470..509
FT                   /note= "BGP (391-430)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
```

```
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 17; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   509 AA;


  Query Match             100.0%;  Score 45;  DB 16;  Length 509;
  Best Local Similarity   100.0%;  Pred. No. 0.34;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGANLNL 9
            |||||||||
Db      396 ylsganlnl 404


RESULT  12
R77438
ID   R77438 standard; Protein; 511 AA.
XX
AC   R77438;
XX
DT   19-JAN-1996  (first entry)
XX
DE   BGP (1-314)/CEA (490-642)/BGP (387-430) chimaeric protein.
XX
KW   Primer; amplify; polymerse chain reaction; PCR; human;
KW   biliary glycoprotein; BGP; membrane-bound; carcinoembryonic antigen;
KW   CEA; chimaeric protein; PR1A3 epitope; anti-PR1A3 antibody;
KW   colorectal carcinoma; monoclonal antibody.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..314
FT                   /note= "BGP (1-314)"
FT   Protein         315..467
FT                   /note= "CEA (490-642)"
FT   Protein         468..511
FT                   /note= "BGP (387-430)"
XX
PN   WO9506067-A1.
XX
PD   02-MAR-1995.
XX
PF   19-AUG-1994;   94WO-GB01816.
XX
PR   21-AUG-1993;   93GB-0017423.
```

```
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Bates PA,  Bodmer WF,  Durbin H,  Snary D,  Stewart LMD;
PI   Young S;
XX
DR   WPI; 1995-106813/14.
XX
PT   New molecules which bind carcinoembryonic antigen - used for the
PT   diagnosis and treatment of colorectal carcinoma and for isolation
PT   and purifications.
XX
PS   Claim 18; ; 67pp; English.
XX
CC   The sequences given in R77435-38 are chimaeric proteins comprising
CC   portions of human biliary glycoprotein (BGP) and the human membrane-
CC   bound carcinoembryonic antigen (CEA).  These chimaeric proteins were
CC   used to identify the PR1A3 epitope.  The PR1A3 epitope was found to be
CC   an epitope of CEA within the B3-GPI region.  Chimaeras in which the CEA
CC   is no longer membrane bound did not react with anti-PR1A3 antibodies
CC   indicating that the PR1A3 epitope is not present in non-membrane bound
CC   hybrid proteins.  Antibodies which recognise the PR1A3 epitope are used
CC   in the detection of well and poorly differentiated colorectal carcinomas.
CC   The isolation of the specific PR1A3 epitope allows the development of
CC   monoclonal antibodies specific for colorectal carcinoma.  They can be
CC   used in the study, isolation and purification of molecules to which they
CC   specifically bind and the imaging and treatment of cells exhibiting the
CC   molecules.
XX
SQ   Sequence   511 AA;


  Query Match             100.0%;  Score 45;  DB 16;  Length 511;
  Best Local Similarity   100.0%;  Pred. No. 0.34;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANLNL 9
           |||||||||
Db     396 ylsganlnl 404


RESULT  13
R60619
ID   R60619 standard; Protein; 642 AA.
XX
AC   R60619;
XX
DT   10-MAY-1995  (first entry)
XX
DE   Carcinoembryonic antigen glycoprotein.
XX
KW   Carcinoembryonic antigen; CEA; neoplastic diseases.
XX
OS   Homo sapiens.
XX
PN   EP618292-A.
XX
PD   05-OCT-1994.
XX
PF   15-MAR-1994;   94EP-0103986.
XX
PR   25-MAR-1993;   93EP-0810214.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Mach J,  Pelegrin A,  Terskikh A;
XX
DR   WPI; 1994-304461/38.
DR   N-PSDB; Q71567.
XX
PT   Carcinoembryonic antigen (CEA) derivs - useful as reagents in
PT   immunoassay for diagnosis of neoplastic diseases
XX
```

```
PS   Claim 2; Page 15; 30pp; English.
XX
CC   Q71567 is the DNA sequence which encodes carcinoembryonic antigen (CEA)
CC   R60619. CEA is free from cross-reactive CEA-like antigens, it is
CC   antigenically indistinguishable from the solution form of CEA shed from
CC   tumour cells, and it is devoid of ethanolamine. R60619 can be used in a
CC   reagent composition for detecting neoplastic diseases in biological
CC   samples, or in an immunoassay process where it can specifically detect
CC   the presence of tumour cells in a biological sample e.g. blood.
XX
SQ   Sequence   642 AA;

  Query Match             100.0%;  Score 45;  DB 15;  Length 642;
  Best Local Similarity   100.0%;  Pred. No. 0.44;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANLNL 9
           |||||||||
Db     571 ylsganlnl 579


RESULT  14
R98519
ID   R98519 standard; Protein; 663 AA.
XX
AC   R98519;
XX
DT   13-NOV-1996  (first entry)
XX
DE   Immunogenic carcinoembryonic antigen.
XX
KW   Carcinoembryonic antigen; immunogen; breast cancer; lung cancer;
KW   colon cancer; therapy; immunotherapy; vaccine; baculovirus; vector;
KW   Spodoptera frugiperda; insect; pA9080 AcNPV-CEA.
XX
OS   Chimeric Autographa californica nuclear polyhedrosis virus;
OS   Chimeric Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..18
FT                   /label= Sig_peptide
FT                   /note= "AcNPV 61k protein signal peptide"
FT   Protein         18..663
FT                   /label= Mat_protein
FT                   /note= "amino acids 1-3 of the mature protein
FT                          are derived from the baculovirus vector"
XX
PN   WO9532286-A2.
XX
PD   30-NOV-1995.
XX
PF   19-MAY-1995;   95WO-US06373.
XX
PR   20-MAY-1994;   94US-0246981.
XX
PA   (MICR-) MICROGENESYS INC.
XX
PI   Hackett C,  Smith G,  Volvovitz F;
XX
DR   WPI; 1996-020581/02.
DR   N-PSDB; T36495.
XX
PT   Immunogenic carcinoembryonic antigen produced using insect cell
PT   baculovirus expression system - useful in cancer therapy
XX
PS   Claim 9; Page 50-53; 61pp; English.
XX
CC   A recombinant, soluble, immunogenic carcinoembryonic antigen
CC   (rCEA) (R98519) is encoded by vector pA9080 AcNPV-CEA (see also
CC   T36495) in which a modified human CEA gene is joined to a
CC   baculovirus signal sequence under control of a polyhedrin
CC   promoter.  The baculovirus signal peptide directs translation of
```

```
CC   rCEA into the insect cell glycosylation pathway.  rCEA can be
CC   produced at high levels in Sf900+ insect cells grown in serum-free
CC   media, and isolated to a purity of over 95%.  It is used as an
CC   immunogen in humans to protect against cancer, partic. breast,
CC   lung or colon cancer.
XX
SQ   Sequence   663 AA;

  Query Match              100.0%;  Score 45;  DB 17;  Length 663;
  Best Local Similarity   100.0%;  Pred. No. 0.46;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 YLSGANLNL 9
             |||||||||
Db       592 ylsganlnl 600


RESULT 15
P81229
ID   P81229 standard; protein; 698 AA.
XX
AC   P81229;
XX
DT   17-NOV-1990  (first entry)
XX
DE   Carcinoembryonic antigen.
XX
KW   Carcinoembryonic antigen; antibody; tumor diagnosis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..30
FT                   /label=signal peptide
FT   Protein         31..668
FT                   /label=mature human CEA
XX
PN   JP63177794-A.
XX
PD   21-JUL-1988.
XX
PF   14-JAN-1987;   87JP-0006851.
XX
PR   14-JAN-1987;   87JP-0006851.
XX
PA   (SUNR ) SUNTORY LTD.
XX
DR   WPI; 1988-245625/35.
DR   N-PSDB; N81611.
XX
PT   Human carcinoembryonic antigen -
PT   used to produce antibodies and detect tumor tissue without
PT   reacting to CEA-related antibody.
XX
PS   Disclosure; ; p; Japanese.
XX
CC   The human carcinoembryonic antigen (CEA) protein sequence is
CC   presented.  Antibodies raised against human CEA are used
CC   to detect tumor tissue (colon carcinoma). See also N81611.
XX
SQ   Sequence   698 AA;

  Query Match              100.0%;  Score 45;  DB 9;  Length 698;
  Best Local Similarity   100.0%;  Pred. No. 0.49;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 YLSGANLNL 9
             |||||||||
Db       601 ylsganlnl 609
```

```
Search completed: December 16, 2000, 00:51:13
Job time: 18773 sec
```

```
                    GenCore version 4.5
             Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 03:07:25 ; Search time 68.03 Seconds
                                               (without alignments)
                                               4.227 Million cell updates/sec

Title:           US-09-529-121-1
Perfect score:   45
Sequence:        1 YLSGANLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters:      213

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       SwissProt_39:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                   %
Result           Query
  No.   Score    Match Length DB  ID                          Description
-----------------------------------------------------------------------------
     1     17    37.8      8  1  CPD1_ENTFA               P13269 enterococcu
     2     16    35.6      8  1  FAR8_CALVO               P41863 calliphora
     3     13    28.9      9  1  ISOT_CYPCA               P42993 cyprinus ca
     4     13    28.9      9  1  OXYT_RAJCL               P42994 raja clavat
     5     13    28.9      9  1  THYF_PIG                 P01255 sus scrofa
     6     12    26.7      7  1  IGAO_DACDE               P06294 dactylium d
     7     12    26.7      9  1  OXYA_SQUAC               P42999 squalus aca
     8     12    26.7      9  1  PGLR_DIAAB               P81179 diaprepes a
     9     11    24.4      4  1  FAR3_HIRME               P42562 hirudo medi
    10     11    24.4      5  1  PRCT_PERAM               P01373 periplaneta
    11     11    24.4      6  1  TMOF_SARBU               P41495 sarcophaga
    12     11    24.4      7  1  FAR2_ASCSU               P31890 ascaris suu
    13     11    24.4      7  1  GFRP_MOUSE               P99025 mus musculu
    14     11    24.4      8  1  CAD1_ENTFA               P13268 enterococcu
    15     11    24.4      8  1  LCK8_LEUMA               P19990 leucophaea
    16     11    24.4      8  1  LPMS_STAEP               P23211 staphylococ
    17     11    24.4      9  1  LMIP_LOCMI               P31799 locusta mig
    18     11    24.4      9  1  MOSF_CLYJA               P19853 clypeaster
    19     11    24.4      9  1  OXYT_RABIT               P32878 oryctolagus
    20     11    24.4      9  1  TAL3_PICJA               P17441 pichia jadi
    21     10    22.2      5  1  TPIS_CANFA               P54714 canis famil
    22     10    22.2      5  1  UXA4_CHLTR               P38005 chlamydia t
    23     10    22.2      6  1  CIP2_MYTED               P13737 mytilus edu
    24     10    22.2      7  1  ALL2_CARMA               P81805 carcinus ma
    25     10    22.2      7  1  ALL3_CARMA               P81806 carcinus ma
    26     10    22.2      7  1  ALL4_CARMA               P81807 carcinus ma
    27     10    22.2      7  1  ALL5_CARMA               P81808 carcinus ma
    28     10    22.2      7  1  FAR1_PROCL               P38499 procambarus
    29     10    22.2      7  1  UN06_PINPS               P81675 pinus pinas
    30     10    22.2      8  1  AKH_MELML                P25423 melolontha
    31     10    22.2      8  1  AL12_CARMA               P81815 carcinus ma
    32     10    22.2      8  1  AL17_CARMA               P81820 carcinus ma
    33     10    22.2      8  1  ALL1_CYDPO               P82152 cydia pomon
    34     10    22.2      8  1  ALL6_CYDPO               P82157 cydia pomon
    35     10    22.2      8  1  ALL7_CARMA               P81809 carcinus ma
    36     10    22.2      8  1  ALL8_CARMA               P81811 carcinus ma
    37     10    22.2      8  1  ALL9_CARMA               P81812 carcinus ma
    38     10    22.2      8  1  FAR4_HOMAM               P41487 homarus ame
    39     10    22.2      8  1  GLUR_HUMAN               P02729 homo sapien
    40     10    22.2      8  1  HTF_TENMO                P25419 tenebrio mo
    41     10    22.2      8  1  LCK3_LEUMA               P21142 leucophaea
    42     10    22.2      8  1  LCK5_LEUMA               P19987 leucophaea
    43     10    22.2      8  1  RPCH_PANBO               P08939 pandalus bo
    44     10    22.2      8  1  UF06_MOUSE               P38644 mus musculu
    45     10    22.2      8  1  VGLG_HSV2B               P81780 herpes simp
    46     10    22.2      9  1  AL10_CARMA               P81813 carcinus ma
    47     10    22.2      9  1  AL11_CARMA               P81814 carcinus ma
    48     10    22.2      9  1  DSIP_RABIT               P01158 oryctolagus
    49     10    22.2      9  1  FAR5_ASCSU               P43170 ascaris suu
    50     10    22.2      9  1  FAR6_CALVO               P41861 calliphora
    51     10    22.2      9  1  FARP_CALSI               P38495 callinectes
    52     10    22.2      9  1  FIBB_MACFU               P19345 macaca fusc
    53     10    22.2      9  1  MOSH_CLYJA               P19852 clypeaster
    54     10    22.2      9  1  OXYA_SCYCA               P42996 scyliorhinu
    55     10    22.2      9  1  SAMP_MUSCA               P19095 mustelus ca
    56     10    22.2      9  1  TKL1_LOCMI               P16223 locusta mig
    57     10    22.2      9  1  TRP4_LEUMA               P81736 leucophaea
    58     10    22.2      9  1  ULAH_HUMAN               P31934 homo sapien
    59     10    22.2      9  1  UPA6_HUMAN               P30092 homo sapien
    60      9    20.0      4  1  FAR4_HIRME               P42563 hirudo medi
    61      9    20.0      5  1  AL14_CARMA               P81817 carcinus ma
    62      9    20.0      5  1  FARP_ARTTR               P41853 artioposthi
    63      9    20.0      7  1  CHOX_ALCSP               P16101 alcaligenes
    64      9    20.0      7  1  FAR3_HAECO               P81298 haemonchus
    65      9    20.0      7  1  FAR3_PANRE               P41874 panagrellus
    66      9    20.0      7  1  FAR5_HIRME               P42564 hirudo medi
    67      9    20.0      7  1  LANC_CARUI               P36960 carnobacter
    68      9    20.0      7  1  UF03_MOUSE               P38641 mus musculu
    69      9    20.0      8  1  AL15_CARMA               P81818 carcinus ma
    70      9    20.0      8  1  AL16_CARMA               P81819 carcinus ma
    71      9    20.0      8  1  AL18_CARMA               P81821 carcinus ma
    72      9    20.0      8  1  ALL3_CYDPO               P82154 cydia pomon
    73      9    20.0      8  1  ALL4_CALVO               P41840 calliphora
    74      9    20.0      8  1  ALL4_CYDPO               P82155 cydia pomon
    75      9    20.0      8  1  ALL5_CYDPO               P82156 cydia pomon


                              ALIGNMENTS


RESULT   1
CPD1_ENTFA
ID   CPD1_ENTFA     STANDARD;      PRT;     8 AA.
AC   P13269;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   SEX PHEROMONE CPD1.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 85040388.
RA   Suzuki A., Mori M., Sagakami Y., Isogai A., Fujino M., Kitada C.,
RA   Craig R.A., Clewell D.B.;
RT   "Isolation and structure of bacterial sex pheromone, cPD1.";
RL   Science 226:849-850(1984).
CC   -!- FUNCTION: CPD1 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE
CC       BACTERIOCIN PLASMID PPD1.
KW   Pheromone.
SQ   SEQUENCE   8 AA;  913 MW;  8665B729C682C729 CRC64;

  Query Match             37.8%;  Score 17;  DB 1;  Length 8;
  Best Local Similarity   75.0%;  Pred. No. 8.8e+04;
```

```
  Matches     3;  Conservative     1;  Mismatches     0;  Indels     0;  Gaps     0;

Qy        1 YLSG 4
            :|||
Db        5 FLSG 8


RESULT   2
FAR8_CALVO
ID   FAR8_CALVO     STANDARD;      PRT;     8 AA.
AC   P41863;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CALLIFMRFAMIDE 8.
OS   Calliphora vomitoria (Blue blowfly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Calliphoridae; Calliphora.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=THORACIC GANGLION;
RX   MEDLINE; 92196111.
RA   Duve H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA   Rehfeld J.F., Thorpe A.;
RT   "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT   neuropeptides (designated calliFMRFamides) from the blowfly
RT   Calliphora vomitoria.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
DR   PIR; H41978; H41978.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  957 MW;  72D40699CAA44DD8 CRC64;


  Query Match             35.6%;  Score 16;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches     3;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy        4 GAN 6
            |||
Db        1 GAN 3


RESULT   3
ISOT_CYPCA
ID   ISOT_CYPCA     STANDARD;      PRT;     9 AA.
AC   P42993;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   ISOTOCIN.
OS   Cyprinus carpio (Common carp).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PITUITARY;
RA   Acher R., Chauvet J., Chauvet M.-T., Crepy D.;
RT   "Characterization of neurohypophyseal hormones from a fresh water bony
RT   fish, the carp (Cyprinus carpio). Comparison with hormones from sea
RT   water bony fishs.";
RL   Comp. Biochem. Physiol. 14:245-254(1965).
CC   -!- FUNCTION: ANTIDIURETIC HORMONE.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   PIR; A61364; A61364.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  969 MW;  17FF476EB455B04B CRC64;


  Query Match             28.9%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches     2;  Conservative     1;  Mismatches     0;  Indels     0;  Gaps     0;

Qy        1 YLS 3
            |:|
Db        2 YIS 4


RESULT   4
OXYT_RAJCL
ID   OXYT_RAJCL     STANDARD;      PRT;     9 AA.
AC   P42994;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   GLUMITOCIN.
OS   Raja clavata (Thornback ray).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Squalea; Hypnosqualea; Pristiorajea; Batoidea;
OC   Rajiformes; Rajidae; Raja.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 66123415.
RA   Acher R., Chauvet J., Chauvet M.-T., Crepy D.;
RT   "Phylogeny of neurophyophyseal peptides: isolation of a new hormone,
RT   glumitocin (Ser 4-Gln 8-ocytocin) present in a cartilaginous fish,
RT   the ray (Raia clavata).";
RL   Biochim. Biophys. Acta 107:393-396(1965).
CC   -!- FUNCTION: ANTIDIURETIC HORMONE.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Amidation.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  984 MW;  17E9C76EB455B04B CRC64;


  Query Match             28.9%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches     2;  Conservative     1;  Mismatches     0;  Indels     0;  Gaps     0;

Qy        1 YLS 3
            |:|
Db        2 YIS 4


RESULT   5
THYF_PIG
ID   THYF_PIG       STANDARD;      PRT;     9 AA.
AC   P01255;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   21-JUL-1986 (Rel. 01, Last annotation update)
DE   THYMIC FACTOR.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 78026571.
RA   Pleau J.-M., Dardenne M., Blouquit Y., Bach J.-F.;
RT   "Structural study of circulating thymic factor: a peptide isolated
RT   from pig serum. II. Amino acid sequence.";
RL   J. Biol. Chem. 252:8045-8047(1977).
CC   -!- MISCELLANEOUS: THE BIOLOGICAL SOURCE(S) AND PHYSIOLOGICAL
```

```
CC        ACTIVITIES OF THIS PEPTIDE HAVE NOT BEEN DETERMINED.
DR    PIR; A01523; YFPG.
FT    MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
SQ    SEQUENCE   9 AA;  876 MW;  D500B87866C5B33D CRC64;


  Query Match             28.9%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GAN 6
           |:|
Db       7 GSN 9


RESULT   6
IGAO_DACDE
ID    IGAO_DACDE     STANDARD;      PRT;     7 AA.
AC    P06294;
DT    01-JAN-1988 (Rel. 06, Created)
DT    01-JAN-1988 (Rel. 06, Last sequence update)
DT    01-OCT-1994 (Rel. 30, Last annotation update)
DE    GALACTOSE OXIDASE INHIBITOR.
OS    Dactylium dendroides (Cladobotryum dendroides).
OC    Eukaryota; Fungi; Ascomycota; Hypocreales; Hypocreaceae; Hypomyces.
RN    [1]
RP    SEQUENCE.
RA    Avigad G., Markus Z.;
RT    "Identification of a peptide inhibitor of galactose oxidase from
RT    Dactylium dendroides.";
RL    Fed. Proc. 31:447-447(1972).
CC    -!- FUNCTION: BINDS ONE COPPER ION PER MOLECULE BUT DOES NOT BIND THE
CC        GALACTOSE OXIDASE APOENZYME. IT MAY INACTIVATE THE ENZYME BY
CC        BINDING TO ITS PROSTHETIC COPPER GROUP.
DR    PIR; A01341; XEYDGD.
KW    Copper; Metalloenzyme inhibitor.
SQ    SEQUENCE   7 AA;  706 MW;  75BB01A456D87DB0 CRC64;


  Query Match             26.7%;  Score 12;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGAN 6
            :| |
Db       1 AGQN 4


RESULT   7
OXYA_SQUAC
ID    OXYA_SQUAC     STANDARD;      PRT;     9 AA.
AC    P42999;
DT    01-NOV-1995 (Rel. 32, Created)
DT    01-NOV-1995 (Rel. 32, Last sequence update)
DT    01-NOV-1995 (Rel. 32, Last annotation update)
DE    ASPARTOCIN (ASPARGTOCIN).
OS    Squalus acanthias (Spiny dogfish).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC    Elasmobranchii; Squalea; Squaloidei; Squalidae; Squalus.
RN    [1]
RP    SEQUENCE.
RX    MEDLINE; 73031727.
RA    Acher R., Chauvet J., Chauvet M.-T.;
RT    "Phylogeny of the neurohypophysial hormones. Two new active peptides
RT    isolated from a cartilaginous fish, Squalus acanthias.";
RL    Eur. J. Biochem. 29:12-19(1972).
RN    [2]
RP    SEQUENCE.
RX    MEDLINE; 72128038.
RA    Acher R., Chauvet J., Chauvet M.-T., Fontaine M.;
RT    "Identification of 2 new neurohypophyseal hormones, valitocin (Val8-
RT    oxytocin) and aspartocin (Asn4-oxytocin) in a selachian fish, the
RT    spiny dog-fish (Squalus acanthias).";
```

```
RL    C. R. Acad. Sci., D, Sci. Nat. 274:313-316(1972).
CC    -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR    INTERPRO; IPR000981; -.
DR    PFAM; PF00220; hormone4; 1.
DR    PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW    Hormone; Amidation.
FT    DISULFID      1      6
FT    MOD_RES       9      9       AMIDATION.
SQ    SEQUENCE   9 AA;  996 MW;  17F8376EB444404B CRC64;


  Query Match             26.7%;  Score 12;  DB 1;  Length 9;
  Best Local Similarity   28.6%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 YLSGANL 7
           |::   |
Db       2 YINNCPL 8


RESULT   8
PGLR_DIAAB
ID    PGLR_DIAAB     STANDARD;      PRT;     9 AA.
AC    P81179;
DT    15-JUL-1998 (Rel. 36, Created)
DT    15-JUL-1998 (Rel. 36, Last sequence update)
DT    30-MAY-2000 (Rel. 39, Last annotation update)
DE    ENDO-POLYGALACTURONASE (PG) (EC 3.2.1.15) (FRAGMENT).
OS    Diaprepes abbreviatus (Sugarcane rootstalk borer weevil).
OC    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC    Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC    Cucujiformia; Curculionidae; Entiminae; Entimini; Diaprepes.
RN    [1]
RP    SEQUENCE.
RC    TISSUE=LARVAL GUT;
RA    Doostdar H., McCollum T.G., Mayer R.T.;
RT    "Purification and characterization of an endo-polygalacturonase from
RT    the gut of West Indies sugarcane rootstalk borer weevil (Diaprepes
RT    abbreviatus L.) larvae.";
RL    Comp. Biochem. Physiol. 118B:861-867(1997).
CC    -!- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-ALPHA-D-
CC        GALACTOSIDURONIC LINKAGES IN PECTATE AND OTHER GALACTURONANS.
CC    -!- INDUCTION: INHIBITED BY CITRUS PGIP.
CC    -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS PROTEIN IS:
CC        9.4, ITS MW IS: 44.5 KDA.
CC    -!- SIMILARITY: WEAK TO OTHER POLYGALACTURONASES.
KW    Hydrolase; Glycosidase; Cell wall.
FT    NON_TER       9      9
SQ    SEQUENCE   9 AA;  1041 MW;  1F49087042DB41BB CRC64;


  Query Match             26.7%;  Score 12;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |: |
Db       4 YVIG 7


RESULT   9
FAR3_HIRME
ID    FAR3_HIRME     STANDARD;      PRT;     4 AA.
AC    P42562;
DT    01-NOV-1995 (Rel. 32, Created)
DT    01-NOV-1995 (Rel. 32, Last sequence update)
DT    01-NOV-1995 (Rel. 32, Last annotation update)
DE    FMRFAMIDE-LIKE NEUROPEPTIDE YLRF-AMIDE.
OS    Hirudo medicinalis (Medicinal leech).
OC    Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;
OC    Arynchobdellida; Hiruliniformes; Hirudinidae; Hirudo.
RN    [1]
RP    SEQUENCE.
```

```
RX   MEDLINE; 92195954.
RA   Evans B.D., Pohl J., Kartsonis M.A., Calabrese R.L.;
RT   "Identification of RFamide neuropeptides in the medicinal leech.";
RL   Peptides 12:897-908(1991).
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES      4      4       AMIDATION.
SQ   SEQUENCE   4 AA;  598 MW;  69D4073B30000000 CRC64;


  Query Match             24.4%;  Score 11;  DB 1;  Length 4;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YL 2
            ||
Db        1 YL 2


RESULT  10
PRCT_PERAM
ID   PRCT_PERAM     STANDARD;      PRT;     5 AA.
AC   P01373;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   PROCTOLIN.
OS   Periplaneta americana (American cockroach),
OS   Limulus polyphemus (Atlantic horseshoe crab), and
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blattoidea; Blattidae; Periplaneta.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=P.AMERICANA;
RX   MEDLINE; 76074708.
RA   Starratt A.N., Brown B.E.;
RT   "Structure of the pentapeptide proctolin, a proposed neurotransmitter
RT   in insects.";
RL   Life Sci. 17:1253-1256(1975).
RN   [2]
RP   BIOLOGICAL SOURCE.
RC   SPECIES=P.AMERICANA;
RX   MEDLINE; 81225865.
RA   O'Shea M., Adams M.E.;
RT   "Pentapeptide (proctolin) associated with an identified neuron.";
RL   Science 213:567-569(1981).
RN   [3]
RP   SEQUENCE.
RC   SPECIES=L.POLYPHEMUS;
RX   MEDLINE; 90287800.
RA   Groome J.R., Tillinghast E.K., Townley M.A., Vetrovs A.,
RA   Watson W.H. III, Hunt D.F., Griffin P.R., Alexander J.E.,
RA   Shabanowitz J.;
RT   "Identification of proctolin in the central nervous system of the
RT   horseshoe crab, Limulus polyphemus.";
RL   Peptides 11:205-211(1990).
RN   [4]
RP   SEQUENCE.
RC   SPECIES=C.MAENAS;
RX   MEDLINE; 86232789.
RA   Stangier J., Dircksen H., Keller R.;
RT   "Identification and immunocytochemical localization of proctolin in
RT   pericardial organs of the shore crab, Carcinus maenas.";
RL   Peptides 7:67-72(1986).
CC   -!- FUNCTION: STIMULATES CARDIAC OUTPUT AND HINDGUT MOTILITY,
CC       MODULATES VISCERAL AND SKELETAL MUSCLE IN MANY ARTHROPODS.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE LATERAL WHITE NEURONS AND IN
CC       THE CRAB PERICARDIAL ORGANS.
DR   PIR; A01644; HOROHA.
DR   PIR; A60411; A60411.
```

```
KW   Neuropeptide.
SQ   SEQUENCE   5 AA;  649 MW;  71B7673B44600000 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YL 2
            ||
Db        2 YL 3


RESULT  11
TMOF_SARBU
ID   TMOF_SARBU     STANDARD;      PRT;     6 AA.
AC   P41495;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   TRYPSIN-MODULATING OOSTATIC FACTOR (TMOF).
OS   Sarcophaga bullata (Grey flesh fly) (Neobellieria bullata).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Sarcophagidae; Sarcophaga.
RN   [1]
RP   SEQUENCE, AND SYNTHESIS.
RC   TISSUE=OVARY;
RX   MEDLINE; 94211930.
RA   Bylemans D., Borovsky D., Hunt D.F., Shabanowitz J., Grauwels L.,
RA   de Loof A.;
RT   "Sequencing and characterization of trypsin modulating oostatic
RT   factor (TMOF) from the ovaries of the grey fleshfly, Neobellieria
RT   (Sarcophaga) bullata.";
RL   Regul. Pept. 50:61-72(1994).
CC   -!- FUNCTION: HAS AN OOSTATIC ACTIVITY. INHIBITS TRYPSIN BIOSYNTHESIS
CC       IN THE MIDGUT WHICH INDIRECTLY REDUCES THE VITELLOGENIN
CC       CONCENTRATION IN THE HEMOLYMPH RESULTING IN INHIBITION OF OOCYTE
CC       DEVELOPMENT.
CC   -!- DEVELOPMENTAL STAGE: SYNTHESIZED AND RELEASED FROM FOLLICULAR
CC       EPITHELIUM AFTER A BLOOD MEAL.
KW   Hormone.
SQ   SEQUENCE   6 AA;  695 MW;  61E72451B7642000 CRC64;


  Query Match             24.4%;  Score 11;  DB 1;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 NLN 8
             ||:
Db        4 NLH 6


RESULT  12
FAR2_ASCSU
ID   FAR2_ASCSU     STANDARD;      PRT;     7 AA.
AC   P31890;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   FMRFAMIDE-LIKE NEUROPEPTIDE AF2.
OS   Ascaris suum (Pig roundworm) (Ascaris lumbricoides), and
OS   Panagrellus redivivus.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC   Ascarididae; Ascaris.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=A.SUUM;
RX   MEDLINE; 93324431.
RA   Cowden C., Stretton A.O.W.;
RT   "AF2, an Ascaris neuropeptide: isolation, sequence, and bioactivity.";
RL   Peptides 14:423-430(1993).
```

```
RN   [2]
RP   SEQUENCE.
RC   SPECIES=P.REDIVIVUS;
RX   MEDLINE; 95060998.
RA   Maule A.G., Shaw C., Bowman J.W.;
RT   "The FMRFamide-like neuropeptide AF2 (Ascaris suum) is present in the
RT   free-living nematode, Panagrellus redivivus (Nematoda, Rhabditida).";
RL   Parasitology 109:351-356(1994).
CC   -!- FUNCTION: HAS EFFECTS ON MUSCLE TENSION.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE NERVE CORDS AND A VARIETY OF
CC       GANGLIA PARTICULARLY IN THE ANTERIOR REGIONS.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  992 MW;  69D4073B5B11E350 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YL 2
              ||
Db          4 YL 5


RESULT  13
GFRP_MOUSE
ID   GFRP_MOUSE     STANDARD;      PRT;     7 AA.
AC   P99025;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   GTP CYCLOHYDROLASE I FEEDBACK REGULATORY PROTEIN (P35) (FRAGMENT).
GN   GFRP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LIVER;
RA   Sanchez J.-C., Rouge V., Frutiger S., Hughes G., Yan J.X.,
RA   Hoogland C., Appel R.D., Binz P.-A., Hochstrasser D.F.,
RA   Cowthorne M.;
RL   Submitted (AUG-1998) to the SWISS-PROT data bank.
CC   -!- FUNCTION: MEDIATES TETRAHYDROBIOPTERIN INHIBITION OF GTP
CC       CYCLOHYDROLASE I. THIS INHIBITION IS REVERSED BY L-PHENYLALANINE
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
DR   SWISS-2DPAGE; P99025; MOUSE.
FT   INIT_MET      0      0
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  806 MW;  71B5B057273B4700 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 YL 2
              ||
Db          2 YL 3


RESULT  14
CAD1_ENTFA
ID   CAD1_ENTFA     STANDARD;      PRT;     8 AA.
AC   P13268;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   SEX PHEROMONE CAD1.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 85051889.
RA   Mori M., Sagakami Y., Narita M., Isogai A., Fujino M., Kitada C.,
RA   Craig R.A., Clewell D.B., Suzuki A.;
RT   "Isolation and structure of the bacterial sex pheromone, cAD1, that
RT   induces plasmid transfer in Streptococcus faecalis.";
RL   FEBS Lett. 178:97-100(1984).
CC   -!- FUNCTION: CAD1 IS INVOLVED IN THE CONJUGATIVE TRANSFER OF THE
CC       HEMOLYSIN PLASMID PAD1.
KW   Pheromone.
SQ   SEQUENCE   8 AA;  819 MW;  047DD732C735B9C7 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LSG 4
              |:|
Db          6 LAG 8


RESULT  15
LCK8_LEUMA
ID   LCK8_LEUMA     STANDARD;      PRT;     8 AA.
AC   P19990;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   LEUCOKININ VIII (L-VIII).
OS   Leucophaea maderae (Madeira cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blaberoidea; Blaberidae; Leucophaea.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=HEAD;
RA   Holman G.M., Cook B.J., Nachman R.J.;
RT   "Isolation, primary structure and synthesis of leucokinins VII and
RT   VIII: the final members of this new family of cephalomyotropic
RT   peptides isolated from head extracts of Leucophaea maderae.";
RL   Comp. Biochem. Physiol. 88C:31-34(1987).
CC   -!- FUNCTION: THIS CEPHALOMYOTROPIC PEPTIDE STIMULATES CONTRACTILE
CC       ACTIVITY OF COCKROACH PROTODEUM (HINDGUT).
CC   -!- SIMILARITY: TO THE OTHER LEUCOKININS.
DR   PIR; JS0318; JS0318.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  902 MW;  736365AB59CAADD8 CRC64;

  Query Match             24.4%;  Score 11;  DB 1;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 GAN 6
              ||:
Db          1 GAD 3


RESULT  16
LPMS_STAEP
ID   LPMS_STAEP     STANDARD;      PRT;     8 AA.
AC   P23211;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   01-MAR-1992 (Rel. 21, Last annotation update)
DE   PROBABLE MSRA LEADER PEPTIDE.
OS   Staphylococcus epidermidis.
```

```
OG   Plasmid pUL5050.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Staphylococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=968;
RX   MEDLINE; 91041730.
RA   Ross J.I., Eady E.A., Cove J.H., Cunliffe W.J., Baumberg S.,
RA   Wootton J.C.;
RT   "Inducible erythromycin resistance in staphylococci is encoded by a
RT   member of the ATP-binding transport super-gene family.";
RL   Mol. Microbiol. 4:1207-1214(1990).
CC   -!- FUNCTION: MAY REGULATE EXPRESSION OF THE ERYTHROMYCIN RESISTANCE
CC       PROTEIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X52085; CAA36303.1; -.
DR   PIR; S11157; LFSAME.
KW   Leader peptide; Plasmid.
SQ   SEQUENCE   8 AA;  937 MW;  FA37340685BDC1A6 CRC64;


  Query Match             24.4%;  Score 11;  DB 1;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        5 ANLNL 9
            |:: |
Db        3 ASMRL 7


RESULT 17
LMIP_LOCMI
ID   LMIP_LOCMI     STANDARD;      PRT;     9 AA.
AC   P31799;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-OCT-1993 (Rel. 27, Last annotation update)
DE   LOCUSTAMYOINHIBITING PEPTIDE (LOM-MIP).
OS   Locusta migratoria (Migratory locust).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC   Acridomorpha; Acridoidea; Acrididae; Locusta.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92179466.
RA   Schoofs L., Holman G.M., Hayes T.K., Nachman R.J., de Loof A.;
RT   "Isolation, identification and synthesis of locustamyoinhibiting
RT   peptide (LOM-MIP), a novel biologically active neuropeptide from
RT   Locusta migratoria.";
RL   Regul. Pept. 36:111-119(1991).
CC   -!- FUNCTION: SUPPRESSES SPONTANEOUS CONTRACTIONS OF THE HINDGUT AND
CC       OVIDUCT.
CC   -!- TISSUE SPECIFICITY: NEURONS LOCATED IN TWO VENTRAL CELL CLUSTERS
CC       IN THE SUBOESOPHAGEAL GANGLION.
DR   PIR; A60065; AKLQIM.
KW   Amidation; Neuropeptide.
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1060 MW;  387D7DD4472AB6C3 CRC64;


  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 NLN 8
```

```
             :||
Db        4 DLN 6


RESULT 18
MOSF_CLYJA
ID   MOSF_CLYJA     STANDARD;      PRT;     9 AA.
AC   P19853;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   [PHE-6]-MOSACT.
OS   Clypeaster japonicus (Sand dollar).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Gnathostomata; Clypeasteroida;
OC   Clypeasteridae; Clypeaster.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=EGG JELLY;
RA   Suzuki N., Kurita M., Yoshino K.I., Kajiura H., Nomura K.,
RA   Yamaguchi M.;
RT   "Purification and structure of mosact and its derivatives from the
RT   egg jelly of the sea urchin Clypeaster japonicus.";
RL   Zool. Sci. 4:649-656(1987).
CC   -!- FUNCTION: STIMULATES SPERM RESPIRATION AND MOTILITY.
DR   PIR; JN0027; JN0027.
SQ   SEQUENCE   9 AA;  924 MW;  93245729CDC5BAB5 CRC64;


  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSG 4
             :| |
Db        6 FLIG 9


RESULT 19
OXYT_RABIT
ID   OXYT_RABIT     STANDARD;      PRT;     9 AA.
AC   P32878; P01188;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   OXYTOCIN (OCYTOCIN).
OS   Oryctolagus cuniculus (Rabbit), Hippopotamus amphibius (Hippopotamus),
OS   Balaenoptera physalus (Finback whale) (Common rorqual),
OS   Tachyglossus aculeatus aculeatus (Australian echidna), and
OS   Hydrolagus colliei (Spotted ratfish) (Pacific ratfish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=RABBIT;
RX   MEDLINE; 72215060.
RA   Chauvet J., Chauvet M.-T., Acher R.;
RT   "Evolution of neurohypophyseal hormones: isolation of active
RT   principles from rabbits and rats.";
RL   Biochimie 53:1099-1104(1971).
RN   [2]
RP   SEQUENCE.
RC   SPECIES=H.AMPHIBIUS;
RX   MEDLINE; 71232719.
RA   Ferguson D.R., Pickering B.T.;
RT   "Arginine and lysine vasopressins in the hippopotamus
RT   neurohypophysis.";
RL   Gen. Comp. Endocrinol. 13:425-429(1969).
RN   [3]
RP   SEQUENCE.
RC   SPECIES=B.PHYSALUS;
RA   Acher R., Chauvet J., Chauvet M.-T.;
RT   "Isolation of finback whale oxytocin and vasopressin.";
```

```
RL   Nature 201:191-192(1964).
RN   [4]
RP   SEQUENCE.
RC   SPECIES=A.ACULEATUS;
RX   MEDLINE; 73223515.
RA   Acher R., Chauvet J., Chauvet M.-T.;
RT   "Neurohypophysial hormones and evolution of tetrapods.";
RL   Nature New Biol. 244:124-126(1973).
RN   [5]
RP   SEQUENCE.
RC   SPECIES=H.COLLIEI;
RX   MEDLINE; 70088110.
RA   Pickering B.T., Heller H.;
RT   "Oxytocin as a neurophyophysial hormone in the holocephalian
RT   elasmobranch fish, Hydrolagus colliei.";
RL   J. Endocrinol. 45:597-606(1969).
CC   -!- FUNCTION: OXYTOCIN CAUSES CONTRACTION OF THE SMOOTH MUSCLE OF THE
CC       UTERUS AND OF THE MAMMARY GLAND.
CC   -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
DR   PIR; A91466; A91466.
DR   PIR; A92774; A92774.
DR   PIR; A93147; A93147.
DR   PIR; A93408; A93408.
DR   PIR; B90667; B90667.
DR   PDB; 1XY1; 15-OCT-90.
DR   PDB; 1XY2; 15-OCT-90.
DR   INTERPRO; IPR000981; -.
DR   PFAM; PF00220; hormone4; 1.
DR   PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW   Hormone; Hypothalamus; Amidation; 3D-structure.
FT   DISULFID      1      6
FT   MOD_RES       9      9       AMIDATION.
SQ   SEQUENCE   9 AA;  1010 MW;  17F8376EB456D04B CRC64;


  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   28.6%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 YLSGANL 7
           |:   |
Db       2 YIQNCPL 8


RESULT  20
TAL3_PICJA
ID   TAL3_PICJA     STANDARD;      PRT;     9 AA.
AC   P17441;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   TRANSALDOLASE III (EC 2.2.1.2) (FRAGMENT).
OS   Pichia jadinii (Yeast) (Candida utilis).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Pichia.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 75145197.
RA   Tsolas O., Sun S.C.;
RT   "Isolation of a peptide containing a histidinyl-cysteinyl sequence
RT   from the active center of transaldolase.";
RL   Arch. Biochem. Biophys. 167:525-533(1975).
CC   -!- FUNCTION: TRANSALDOLASE IS IMPORTANT FOR THE BALANCE OF
CC       METABOLITES IN THE PENTOSE-PHOSPHATE PATHWAY.
CC   -!- CATALYTIC ACTIVITY: SEDOHEPTULOSE 7-PHOSPHATE + D-GLYCERALDEHYDE
CC       3-PHOSPHATE = D-ERYTHROSE 4-PHOSPHATE + D-FRUCTOSE 6-PHOSPHATE.
CC   -!- PATHWAY: NONOXYDATIVE PART OF THE PENTOSE-PHOSPHATE PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE TRANSALDOLASE FAMILY.
DR   PIR; A11497; A11497.
DR   INTERPRO; IPR001585; -.
DR   PROSITE; PS00958; TRANSALDOLASE_2; PARTIAL.
DR   PROSITE; PS01054; TRANSALDOLASE_1; PARTIAL.
KW   Transferase; Pentose shunt.
FT   NON_TER       1      1
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1033 MW;  325A31A44EB1E058 CRC64;


  Query Match             24.4%;  Score 11;  DB 1;  Length 9;
  Best Local Similarity   40.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 GANLN 8
           | : |
Db       2 GIHCN 6


RESULT  21
TPIS_CANFA
ID   TPIS_CANFA     STANDARD;      PRT;     5 AA.
AC   P54714;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   TRIOSEPHOSPHATE ISOMERASE (EC 5.3.1.1) (TIM) (FRAGMENT).
GN   TPI1.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=HEART;
RX   MEDLINE; 98163340.
RA   Dunn M.J., Corbett J.M., Wheeler C.H.;
RT   "HSC-2DPAGE and the two-dimensional gel electrophoresis database of
RT   dog heart proteins.";
RL   Electrophoresis 18:2795-2802(1997).
CC   -!- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE = DIHYDROXY-
CC       ACETONE PHOSPHATE.
CC   -!- PATHWAY: PLAYS AN IMPORTANT ROLE IN SEVERAL METABOLIC PATHWAYS.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.
DR   HSC-2DPAGE; P54714; DOG.
DR   INTERPRO; IPR000652; -.
DR   PROSITE; PS00171; TIM; PARTIAL.
KW   Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;
KW   Pentose shunt.
FT   NON_TER       1      1
FT   NON_TER       5      5
SQ   SEQUENCE   5 AA;  550 MW;  64444862C9A00000 CRC64;


  Query Match             22.2%;  Score 10;  DB 1;  Length 5;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GAN 6
           | |
Db       3 GNN 5


RESULT  22
UXA4_CHLTR
ID   UXA4_CHLTR     STANDARD;      PRT;     5 AA.
AC   P38005;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE FROM ELEMENTARY BODY (FRAGMENT).
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE.
RC   STRAIN=L2/434/BU;
RA   Bini L., Santucci A., Magi B., Marzocchi B., Sanchez-Campillo M.,
RA   Comanducci M., Christianen G., Birkelund S., Vtretou E., Ratti G.,
```

```
RA   Pallini V.;
RL   Submitted (SEP-1994) to the SWISS-PROT data bank.
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 4.5, ITS MW IS: 28 KDA.
FT   NON_TER       5      5
SQ   SEQUENCE   5 AA;  474 MW;  75BAA865AA800000 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 SG 4
              ||
Db          2 SG 3


RESULT  23
CIP2_MYTED
ID   CIP2_MYTED     STANDARD;      PRT;     6 AA.
AC   P13737;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   CONTRACTION-INHIBITING PEPTIDE II (MIP II).
OS   Mytilus edulis (Blue mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytiloidea; Mytilidae; Mytilus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PEDAL GANGLION;
RX   MEDLINE; 88240357.
RA   Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.;
RT   "Structures and actions of Mytilus inhibitory peptides.";
RL   Biochem. Biophys. Res. Commun. 152:1376-1382(1988).
CC   -!- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN
CC       MUSCLES.
CC   -!- SIMILARITY: TO MIP I.
DR   PIR; B27696; B27696.
KW   Hormone; Amidation.
FT   MOD_RES       6      6       AMIDATION.
SQ   SEQUENCE   6 AA;  621 MW;  72C9C6876DD81000 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 GA 5
              ||
Db          1 GA 2


RESULT  24
ALL2_CARMA
ID   ALL2_CARMA     STANDARD;      PRT;     7 AA.
AC   P81805;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 2.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
```

```
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   MOD_RES       7      7       AMIDATION (POTENTIAL).
SQ   SEQUENCE   7 AA;  770 MW;  672879CDCB5DDB70 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSG 4
              |  |
Db          3 YAFG 6


RESULT  25
ALL3_CARMA
ID   ALL3_CARMA     STANDARD;      PRT;     7 AA.
AC   P81806;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 3.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Multigene family.
SQ   SEQUENCE   7 AA;  796 MW;  672879CDCB476B70 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSG 4
              |  |
Db          3 YAFG 6


RESULT  26
ALL4_CARMA
ID   ALL4_CARMA     STANDARD;      PRT;     7 AA.
AC   P81807;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 4.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
```

```
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Multigene family.
SQ   SEQUENCE   7 AA;  782 MW;  672879CDCB476AC0 CRC64;


  Query Match             22.2%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSG 4
            |  |
Db        3 YAFG 6


RESULT  27
ALL5_CARMA
ID   ALL5_CARMA     STANDARD;      PRT;     7 AA.
AC   P81808;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 5.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   MOD_RES       7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  781 MW;  672879CDCB476420 CRC64;


  Query Match             22.2%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YLSG 4
            |  |
Db        3 YAFG 6


RESULT  28
FAR1_PROCL
ID   FAR1_PROCL     STANDARD;      PRT;     7 AA.
AC   P38499;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CARDIOEXCITATORY FMRFAMIDE HOMOLOG NF1.
OS   Procambarus clarkii (Red swamp crayfish).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC   Astacoidea; Cambaridae; Procambarus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=PERICARDIAL ORGANS;
RX   MEDLINE; 93248032.
RA   Mercier A.J., Orchard I., Tebrugge V., Skerrett M.;
RT   "Isolation of two FMRFamide-related peptides from crayfish
RT   pericardial organs.";
RL   Peptides 14:137-143(1993).
```

```
CC   -!- FUNCTION: INCREASES THE RATE AND AMPLITUDE OF SPONTANEOUS
CC       CONTRACTIONS OF SEMI-ISOLATED HEARTS. INCREASES THE AMPLITUDE OF
CC       EXCITATORY POSTSYNAPTIC POTENTIALS IN ABDOMINAL EXTENSOR MUSCLE.
CC   -!- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
CC       FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       7      7       AMIDATION.
SQ   SEQUENCE   7 AA;  966 MW;  69D40729C4540420 CRC64;


  Query Match             22.2%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        6 NLN 8
             | |
Db        1 NRN 3


RESULT  29
UN06_PINPS
ID   UN06_PINPS     STANDARD;      PRT;     7 AA.
AC   P81675;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE OF NEEDLES (N141) (FRAGMENT).
OS   Pinus pinaster (Maritime pine).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Coniferopsida; Coniferales; Pinaceae; Pinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=NEEDLE;
RX   MEDLINE; 99274088.
RA   Costa P., Pionneau C., Bauw G., Dubos C., Bahrman N., Kremer A.,
RA   Frigerio J.-M., Plomion C.;
RT   "Separation and characterization of needle and xylem maritime pine
RT   proteins.";
RL   Electrophoresis 20:1098-1108(1999).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 6.6, ITS MW IS: 25 KDA.
FT   NON_TER       1      1
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  823 MW;  69D76724486B5740 CRC64;


  Query Match             22.2%;  Score 10;  DB 1;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 NL 7
            ||
Db        4 NL 5


RESULT  30
AKH_MELML
ID   AKH_MELML      STANDARD;      PRT;     8 AA.
AC   P25423;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   ADIPOKINETIC HORMONE (AKH).
OS   Melolontha melolontha (Cockchafer),
OS   Geotrupes stercorosus (Dor beetle), and
OS   Pachnoda marginata (Flower beetle).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC   Scarabaeiformia; Scarabaeidae; Melolonthinae; Melolontha.
RN   [1]
RP   SEQUENCE.
RC   SPECIES=M.MELOLONTHA, AND G.STERCOROSUS; TISSUE=CORPORA CARDIACA;
RX   MEDLINE; 91248100.
```

```
RA   Gaede G.;
RT   "A unique charged tyrosine-containing member of the adipokinetic
RT   hormone/red-pigment-concentrating hormone peptide family isolated and
RT   sequenced from two beetle species.";
RL   Biochem. J. 275:671-677(1991).
RN   [2]
RP   SEQUENCE.
RC   SPECIES=P.MARGINATA; TISSUE=CORPORA CARDIACA;
RX   MEDLINE; 92265187.
RA   Gaede G., Lopata A., Kellner R., Rinehart K.L. Jr.;
RT   "Primary structures of neuropeptides isolated from the corpora
RT   cardiaca of various cetonid beetle species determined by
RT   pulsed-liquid phase sequencing and tandem fast atom bombardment mass
RT   spectrometry.";
RL   Biol. Chem. Hoppe-Seyler 373:133-142(1992).
CC   -!- FUNCTION: THIS HORMONE, RELEASED FROM CELLS IN THE CORPORA
CC       CARDIACA AFTER THE BEGINNING OF FLIGHT, CAUSES RELEASE OF
CC       DIGLYCERIDES FROM THE FAT BODY AND THEN STIMULATES THE FLIGHT
CC       MUSCLES TO USE THESE DIGLYCERIDES AS AN ENERGY SOURCE.
CC   -!- SIMILARITY: BELONGS TO THE AKH / HRTH / RPCH FAMILY.
DR   PIR; S15422; S15422.
DR   PIR; S21663; S21663.
DR   INTERPRO; IPR002047; -.
DR   PROSITE; PS00256; AKH; 1.
KW   Neuropeptide; Amidation; Flight.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  1022 MW;  867AB775AB544736 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          7 LN 8
              ||
Db          2 LN 3

RESULT 31
AL12_CARMA
ID   AL12_CARMA     STANDARD;      PRT;     8 AA.
AC   P81815;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 12.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Multigene family.
SQ   SEQUENCE   8 AA;  913 MW;  672879CDCB569AB7 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 YLSG 4
              |  |
Db          4 YAFG 7

RESULT 32
AL17_CARMA
ID   AL17_CARMA     STANDARD;      PRT;     8 AA.
AC   P81820;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 17.
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   MOD_RES       8      8       AMIDATION (POTENTIAL).
SQ   SEQUENCE   8 AA;  858 MW;  C82879D5AB46D865 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 SG 4
              ||
Db          1 SG 2

RESULT 33
ALL1_CYDPO
ID   ALL1_CYDPO     STANDARD;      PRT;     8 AA.
AC   P82152;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYDIASTATIN 1.
OS   Cydia pomonella (Codling moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Tortricoidea; Tortricidae; Olethreutinae; Cydia.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LARVA;
RX   MEDLINE; 98054539.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
RA   Davey M., East P.D., Thorpe A.;
RT   "Lepidopteran peptides of the allatostatin superfamily.";
RL   Peptides 18:1301-1309(1997).
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES       8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  934 MW;  C82879C45B51F775 CRC64;

  Query Match             22.2%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          6 NLNL 9
              |  |
Db          5 NFGL 8
```

```
RESULT  34
ALL6_CYDPO
ID   ALL6_CYDPO     STANDARD;      PRT;     8 AA.
AC   P82157;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYDIASTATIN 6.
OS   Cydia pomonella (Codling moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Tortricoidea; Tortricidae; Olethreutinae; Cydia.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=LARVA;
RX   MEDLINE; 98054539.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
RA   Davey M., East P.D., Thorpe A.;
RT   "Lepidopteran peptides of the allatostatin superfamily.";
RL   Peptides 18:1301-1309(1997).
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation.
FT   MOD_RES      8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  936 MW;  0B2879C45B573767 CRC64;


  Query Match             22.2%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 NLNL 9
           |  |
Db       5 NFGL 8


RESULT  35
ALL7_CARMA
ID   ALL7_CARMA     STANDARD;      PRT;     8 AA.
AC   P81809; P81810; P81804;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CARCINUSTATIN 7 [CONTAINS: CARCINUSTATIN 6; CARCINUSTATIN 1].
OS   Carcinus maenas (Common shore crab) (Green crab).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC   Eubrachyura; Portunoidea; Portunidae; Carcinus.
RN   [1]
RP   SEQUENCE.
RC   TISSUE=CEREBRAL GANGLION, AND THORACIC GANGLION;
RX   MEDLINE; 98121193.
RA   Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA   Thorpe A.;
RT   "Isolation and identification of multiple neuropeptides of the
RT   allatostatin superfamily in the shore crab Carcinus maenas.";
RL   Eur. J. Biochem. 250:727-734(1997).
CC   -!- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC   -!- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW   Neuropeptide; Amidation; Multigene family.
FT   CHAIN        1      8       CARCINUSTATIN 7.
FT   CHAIN        2      8       CARCINUSTATIN 6.
FT   CHAIN        4      8       CARCINUSTATIN 1.
FT   MOD_RES      8      8       AMIDATION.
SQ   SEQUENCE   8 AA;  825 MW;  922879CDCB4775BD CRC64;


  Query Match             22.2%;  Score 10;  DB 1;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 8.8e+04;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           |  |
Db       4 YAFG 7


Search completed: December 16, 2000, 04:23:31
Job time: 4566 sec
```

```
                  GenCore version 4.5
         Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 16, 2000, 03:05:26 ; Search time 111.26 Seconds
                                              (without alignments)
                                              7.553 Million cell updates/sec

Title:           US-09-529-121-1
Perfect score:   45
Sequence:        1 YLSGANLNL 9

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters:      436

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       SPTREMBL_14:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                    SUMMARIES
                   %
Result            Query
  No.    Score   Match Length DB  ID                      Description
 -------------------------------------------------------------------------
    1      17     37.8      8 13  P82079                   P82079 limnodynast
    2      15     33.3      8  3  P87225                  P87225 saccharomyc
    3      15     33.3      8  4  Q9UL56                  Q9ul56 homo sapien
    4      15     33.3      8 13  Q9PS69                   Q9ps69 gallus gall
    5      14     31.1      8  2  Q9X3K1                  Q9x3k1 prochloroco
    6      14     31.1      9  2  Q48686                  Q48686 lactococcus
    7      14     31.1      9  6  Q9TRW2                  Q9trw2 oryctolagus
    8      14     31.1      9  6  Q9TRS0                  Q9trs0 oryctolagus
    9      14     31.1      9  7  Q31415                  Q31415 gallus gall
   10      14     31.1      9 11  O35953                   O35953 mus musculu
   11      13     28.9      8  2  P77556                  P77556 escherichia
   12      13     28.9      8  4  Q9UMH9                  Q9umh9 homo sapien
   13      13     28.9      8 13  Q90498                   Q90498 erythrura g
   14      13     28.9      8 13  Q91098                   Q91098 manorina me
   15      13     28.9      8 13  P82082                   P82082 limnodynast
   16      13     28.9      8 13  P82083                   P82083 limnodynast
   17      13     28.9      9  5  Q9TWD6                  Q9twd6 leptinotars
   18      12     26.7      7 11  Q63480                   Q63480 rattus norv
   19      12     26.7      7 12  Q86871                   Q86871 cauliflower
   20      12     26.7      8  4  Q15889                  Q15889 homo sapien
   21      12     26.7      8  4  Q15901                  Q15901 homo sapien
   22      12     26.7      8  4  Q9Y4J3                  Q9y4j3 homo sapien
   23      12     26.7      8  7  Q95213                  Q95213 oryctolagus
   24      12     26.7      8 12  Q66807                   Q66807 echovirus 2
   25      12     26.7      9  2  Q9R7E8                  Q9r7e8 escherichia
   26      12     26.7      9  2  Q9R635                  Q9r635 chlamydia t
   27      12     26.7      9  4  Q14715                  Q14715 homo sapien
   28      12     26.7      9  4  Q16220                  Q16220 homo sapien
   29      12     26.7      9  4  Q9UCN5                  Q9ucn5 homo sapien
   30      12     26.7      9  5  Q27396                  Q27396 babesia bov
   31      12     26.7      9  6  Q28112                  Q28112 bos taurus
   32      12     26.7      9 12  Q89491                   Q89491 murine minu
   33      11     24.4      7  2  Q47505                  Q47505 escherichia
   34      11     24.4      7 12  Q07624                   Q07624 rous sarcom
   35      11     24.4      7 12  Q9YQ10                   Q9yq10 porcine tra
   36      11     24.4      8  2  Q9RQ57                  Q9rq57 buchnera ap
   37      11     24.4      8  2  Q9RQ49                  Q9rq49 buchnera ap
   38      11     24.4      8  4  Q9UDZ4                  Q9udz4 homo sapien
   39      11     24.4      8  4  Q9UCN4                  Q9ucn4 homo sapien
   40      11     24.4      8  6  Q9XSY1                  Q9xsy1 canis famil
   41      11     24.4      8 11  Q9QVF4                   Q9qvf4 rattus sp.
   42      11     24.4      8 12  Q83332                   Q83332 murine hepa
   43      11     24.4      8 12  Q85562                   Q85562 moloney mur
   44      11     24.4      9  2  Q44001                  Q44001 aeromonas e
   45      11     24.4      9  2  Q44377                  Q44377 aeromonas t
   46      11     24.4      9  2  Q44468                  Q44468 aeromonas v
   47      11     24.4      9  2  Q43928                  Q43928 aeromonas c
   48      11     24.4      9  2  Q9R7H9                  Q9r7h9 haemophilus
   49      11     24.4      9  2  Q9R5M1                  Q9r5m1 staphylococ
   50      11     24.4      9  4  P78484                  P78484 homo sapien
   51      11     24.4      9  5  P82003                  P82003 bombyx mori
   52      11     24.4      9  8  Q31653                  Q31653 anser caeru
   53      11     24.4      9  8  O78337                  O78337 caloglossa
   54      11     24.4      9  8  Q9TLD0                  Q9tld0 bostrychia
   55      11     24.4      9  8  Q9TJ87                  Q9tj87 caloglossa
   56      11     24.4      9  8  Q9TJ85                  Q9tj85 caloglossa
   57      11     24.4      9  8  Q9T389                  Q9t389 caloglossa
   58      11     24.4      9  8  Q9T388                  Q9t388 caloglossa
   59      11     24.4      9  8  Q9T387                  Q9t387 bostrychia
   60      11     24.4      9 10  O81962                   O81962 caloglossa
   61      11     24.4      9 10  O81964                   O81964 caloglossa
   62      11     24.4      9 10  O81966                   O81966 caloglossa
   63      11     24.4      9 10  O81968                   O81968 caloglossa
   64      11     24.4      9 10  O82778                   O82778 caloglossa
   65      11     24.4      9 13  Q92009                   Q92009 gallus gall
   66      10     22.2      5 10  Q99007                   Q99007 hordeum vul
   67      10     22.2      7  5  Q9VYN9                  Q9vyn9 drosophila
   68      10     22.2      7  8  O98866                  O98866 spinacia ol
   69      10     22.2      8  2  P72221                  P72221 pseudomonas
   70      10     22.2      8  2  Q9R7T2                  Q9r7t2 escherichia
   71      10     22.2      8  2  Q9R4M3                  Q9r4m3 enterococcu
   72      10     22.2      8  4  Q16428                  Q16428 homo sapien
   73      10     22.2      8  4  Q9Y4J4                  Q9y4j4 homo sapien
   74      10     22.2      8  4  Q9UMC7                  Q9umc7 homo sapien
   75      10     22.2      8  5  O02032                  O02032 lytechinus

                                    ALIGNMENTS

RESULT   1
P82079
ID   P82079     PRELIMINARY;      PRT;     8 AA.
AC   P82079;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   DYNASTIN 1.
OS   Limnodynastes interioris (Giant banjo frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Myobatrachidae;
OC   Limnodynastes.
```

```
RN   [1]
RP   SEQUENCE, AND MASS SPECTROMETRY.
RC   TISSUE=TIBIAL GLAND;
RA   Raftery M.J., Bradford A.M., Bowie J.H., Wallace J.C., Tyler M.J.;
RT   "Peptides from Australian frogs. The structure of the dynastins from
RT   the banjo frogs Limnodynastes interioras, Limnodynastes dumerilii and
RT   Limnodynastes terraereginae.";
RL   Aust. J. Chem. 46:833-842(1993).
CC   -!- MASS SPECTROMETRY: MW=729; METHOD=FAB.
KW   Amphibian skin.
SQ   SEQUENCE   8 AA;  729 MW;  7C28772865B72728 CRC64;

  Query Match             37.8%;  Score 17;  DB 13;  Length 8;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         2 LSGANL 7
             |||  |
Db         3 LSGLGL 8

RESULT   2
P87225
ID   P87225      PRELIMINARY;      PRT;     8 AA.
AC   P87225;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GIN11 PROTEIN (FRAGMENT).
GN   GIN11.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wedler H., Wedler E., Scharfe M., Wambutt R.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z73169; CAA97518.2; -.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  1019 MW;  4E21A9C449D5B73B CRC64;

  Query Match             33.3%;  Score 15;  DB 3;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 YLS 3
             |||
Db         1 YLS 3

RESULT   3
Q9UL56
ID   Q9UL56      PRELIMINARY;      PRT;     8 AA.
AC   Q9UL56;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MUTANT NADH-CYTOCHROME B5 REDUCTASE (EC 1.6.2.2) (FRAGMENT).
GN   DIA1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fukumaki Y., Higasa K.;
RT   "Two novel mutations in Thai patients with hereditary
RT   methemoglobinemia types I and II: a subtle amino acid change causes
```

```
RT   instability of NADH-cytochrome b5 reductase.";
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF061830; AAF06818.1; -.
KW   Oxidoreductase.
FT   NON_TER       1      1
FT   VARIANT       9      9       -> R.
SQ   SEQUENCE   8 AA;  888 MW;  76C5B73B5051F6D8 CRC64;

  Query Match             33.3%;  Score 15;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 YLS 3
             |||
Db         5 YLS 7

RESULT   4
Q9PS69
ID   Q9PS69      PRELIMINARY;      PRT;     8 AA.
AC   Q9PS69;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   LOW DENSITY LIPOPROTEIN RECEPTOR-RELATED PROTEIN.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92011685.
RA   Stifani S., Barber D.L., Aebersold R., Steyrer E., Shen X., Nimpf J.,
RA   Schneider W.J.;
RT   "The laying hen expresses two different low density lipoprotein
RT   receptor-related proteins.";
RL   J. Biol. Chem. 266:19079-19087(1991).
SQ   SEQUENCE   8 AA;  846 MW;  C007272DD865BAAA CRC64;

  Query Match             33.3%;  Score 15;  DB 13;  Length 8;
  Best Local Similarity   80.0%;  Pred. No. 3e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         3 SGANL 7
             ||| |
Db         3 SGALL 7

RESULT   5
Q9X3K1
ID   Q9X3K1      PRELIMINARY;      PRT;     8 AA.
AC   Q9X3K1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CYTOCHROME B (FRAGMENT).
GN   PETB.
OS   Prochlorococcus sp.
OC   Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC   Prochlorococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Urbach E., Chisholm S.W.;
RT   "Genetic diversity in Prochlorococcus populations flow cytometrically
RT   sorted from the Sargasso Sea and Gulf Stream.";
RL   Limnol. Oceanog. 43:1615-1630(1998).
DR   EMBL; AF070193; AAD23233.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  799 MW;  10376865B72866D3 CRC64;
```

```
  Query Match             31.1%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSG 4
            |||
Db        4 LSG 6


RESULT   6
Q48686
ID   Q48686      PRELIMINARY;      PRT;     9 AA.
AC   Q48686;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   PROMOTER 23 DNA FRAGMENT (FRAGMENT).
OS   Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Lactococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88105390.
RA   van der Vossen J.M., der Lelie D., Venema G.;
RT   "Isolation and characterization of Streptococcus cremoris Wg2-specific
RT   promoters.";
RL   Appl. Environ. Microbiol. 53:2452-2457(1987).
DR   EMBL; M24763; AAA74720.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1080 MW;  5AF3A44AA4469443 CRC64;


  Query Match             31.1%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        6 NLN 8
            |:|
Db        3 NMN 5


RESULT   7
Q9TRW2
ID   Q9TRW2      PRELIMINARY;      PRT;     9 AA.
AC   Q9TRW2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CALDESMON-PHOSPHORYLATION SITE.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 91378498.
RA   Ikebe M., Hornick T.;
RT   "Determination of the phosphorylation sites of smooth muscle caldesmon
RT   by protein kinase C.";
RL   Arch. Biochem. Biophys. 288:538-542(1991).
SQ   SEQUENCE   9 AA;  1018 MW;  8C901B10533735A5 CRC64;


  Query Match             31.1%;  Score 14;  DB 6;  Length 9;
  Best Local Similarity   33.3%;  Pred. No. 3e+05;
  Matches    2;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 GANLNL 9
            |::| :
Db        1 GSSLKI 6


RESULT   8
Q9TRS0
ID   Q9TRS0      PRELIMINARY;      PRT;     9 AA.
AC   Q9TRS0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CALCYCLIN-ASSOCIATED PROTEIN, CAP50=CA2+/PHOSPHOLIPID-BINDING PROTEIN
DE   L-7 FRAGMENT.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92250478.
RA   Tokumitsu H., Mizutani A., Minami H., Kobayashi R., Hidaka H.;
RT   "A calcyclin-associated protein is a newly identified member of the
RT   Ca2+/phospholipid-binding proteins, annexin family.";
RL   J. Biol. Chem. 267:8919-8924(1992).
SQ   SEQUENCE   9 AA;  1010 MW;  64E419C44865B72B CRC64;


  Query Match             31.1%;  Score 14;  DB 6;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSG 4
            |||
Db        3 LSG 5


RESULT   9
Q31415
ID   Q31415      PRELIMINARY;      PRT;     9 AA.
AC   Q31415;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   MHC CLASS I ANTIGEN (FRAGMENT).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kanki T., Kuwasawa N., Sekiya Y., Ichikawa Y.;
RT   "Responsive expression of a MHC class I epitope and genes following
RT   Marek's disease virus infection.";
RL   Submitted (JAN-1992) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; D90399; BAA14395.1; -.
KW   MHC.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  859 MW;  8A55A76455B861B5 CRC64;


  Query Match             31.1%;  Score 14;  DB 7;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 SGAN 6
            :|:|
Db        3 TGSN 6


RESULT  10
O35953
ID   O35953      PRELIMINARY;      PRT;     9 AA.
AC   O35953;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   SODIUM CHANNEL, VOLTAGE-GATED, TYPE VIII, ALPHA POLYPEPTIDE
DE   (TRUNCATED VOLTAGE-GATED SODIUM CHANNEL ALPHA SUBUNIT) (FRAGMENT).
GN   SCN8A.
OS   Mus musculus (Mouse).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIII;
RX   MEDLINE; 97442476.
RA   Plummer N.W., McBurney M.W., Meisler M.H.;
RT   "Alternative splicing of the sodium channel SCN8A predicts a truncated
RT   two-domain protein in fetal brain and non-neuronal cells.";
RL   J. Biol. Chem. 272:24008-24015(1997).
DR   EMBL; U97672; AAB80914.1; -.
DR   MGD; MGI:103169; Scn8a.
KW   Ionic channel.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  898 MW;  22D92865B735B737 CRC64;


  Query Match             31.1%;  Score 14;  DB 11;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           |||
Db       5 LSG 7


RESULT  11
P77556
ID   P77556     PRELIMINARY;      PRT;     8 AA.
AC   P77556;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   TRAY (FRAGMENT).
GN   TRAY.
OS   Escherichia coli.
OG   Plasmid IncFII R1.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ECOR11;
RX   MEDLINE; 96400908.
RA   Boyd E.F., Hill C.W., Rich S.M., Hartl D.L.;
RT   "Mosaic structure of plasmids from natural populations of Escherichia
RT   coli.";
RL   Genetics 143:1091-1100(1996).
DR   EMBL; U50661; AAC44245.1; -.
DR   EMBL; U50650; AAC44234.1; -.
DR   EMBL; U50651; AAC44235.1; -.
DR   EMBL; U50652; AAC44236.1; -.
DR   EMBL; U50653; AAC44237.1; -.
DR   EMBL; U50654; AAC44238.1; -.
DR   EMBL; U50655; AAC44239.1; -.
DR   EMBL; U50656; AAC44240.1; -.
DR   EMBL; U50657; AAC44241.1; -.
DR   EMBL; U50658; AAC44242.1; -.
DR   EMBL; U50659; AAC44243.1; -.
DR   EMBL; U50660; AAC44244.1; -.
KW   Plasmid.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  834 MW;  D335A5B0544735A1 CRC64;


  Query Match             28.9%;  Score 13;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 NLNL 9
           :||:
Db       2 SLNI 5
```

```
RESULT  12
Q9UMH9
ID   Q9UMH9     PRELIMINARY;      PRT;     8 AA.
AC   Q9UMH9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RHCE PROTEIN (FRAGMENT).
GN   RHCE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RA   Matassi G., Cherif-Zahar B., Mouro I., Cartron J.P.;
RT   "Characterization of the recombination hot spot involved in the
RT   genomic rearrangement leading to the hybrid D-CE-D gene in the DVI
RT   phenotype.";
RL   Am. J. Hum. Genet. 60:808-817(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RX   MEDLINE; 90349591.
RA   Cherif-Zahar B., Bloy C., Le Van Kim C., Blanchard D., Bailly P.,
RA   Hermand P., Salmon C., Cartron J.-P., Colin Y.;
RT   "Molecular cloning and protein structure of a human blood group Rh
RT   polypeptide.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:6243-6247(1990).
DR   EMBL; Z97030; CAB09726.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  1049 MW;  C007244691FB5AB1 CRC64;


  Query Match             28.9%;  Score 13;  DB 4;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 NLNL 9
           ::||
Db       4 HMNL 7


RESULT  13
Q90498
ID   Q90498     PRELIMINARY;      PRT;     8 AA.
AC   Q90498;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   MYOGLOBIN (FRAGMENT).
OS   Erythrura gouldiae (Gouldian finch).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Passeridae; Erythrura.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EGG1;
RX   MEDLINE; 98208049.
RA   Heslewood M.M., Elphinstone M.S., Tidemann S.C., Baverstock P.R.;
RT   "Myoglobin intron variation in the Gouldian Finch Erythrura gouldiae
RT   assessed by temperature gradient gel electrophoresis.";
RL   Electrophoresis 19:142-151(1998).
DR   EMBL; U40496; AAC60363.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  890 MW;  C6C1F2C865B046DE CRC64;


  Query Match             28.9%;  Score 13;  DB 13;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy       2 LSGAN 6
           :|| :
Db       3 ISGVH 7


RESULT  14
Q91098
ID   Q91098       PRELIMINARY;      PRT;     8 AA.
AC   Q91098;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   MYOGLOBIN (FRAGMENT).
OS   Manorina melanocephala (noisy miner).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Meliphagidae; Manorina.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D02;
RX   MEDLINE; 98208049.
RA   Heslewood M.M., Elphinstone M.S., Tidemann S.C., Baverstock P.R.;
RT   "Myoglobin intron variation in the Gouldian Finch Erythrura gouldiae
RT   assessed by temperature gradient gel electrophoresis.";
RL   Electrophoresis 19:142-151(1998).
DR   EMBL; U40497; AAC60364.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  890 MW;  C6C1F2C865B046DE CRC64;


  Query Match             28.9%;  Score 13;  DB 13;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LSGAN 6
           :|| :
Db       3 ISGVH 7


RESULT  15
P82082
ID   P82082       PRELIMINARY;      PRT;     8 AA.
AC   P82082;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   DYNASTIN 4.
OS   Limnodynastes salmini (Salmin's-striped frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Myobatrachidae;
OC   Limnodynastes.
RN   [1]
RP   SEQUENCE, AND MASS SPECTROMETRY.
RC   TISSUE=SKIN SECRETION;
RA   Bradford A.M., Raftery M.J., Bowie J.H., Wallace J.C., Tyler M.J.;
RT   "Peptides from Australian frogs. The structure of the dynastins from
RT   Limnodynastes salmini and Fletcherin from Limnodynastes fletcheri.";
RL   Aust. J. Chem. 46:1235-1244(1993).
CC   -!- MASS SPECTROMETRY: MW=772; METHOD=FAB.
KW   Amphibian skin.
SQ   SEQUENCE   8 AA;  772 MW;  7B58772455A2C728 CRC64;


  Query Match             28.9%;  Score 13;  DB 13;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 ANLNL 9
           :|| :
Db       4 SNLGI 8


RESULT  16
P82083
ID   P82083       PRELIMINARY;      PRT;     8 AA.
AC   P82083;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   DYNASTIN 5.
OS   Limnodynastes salmini (Salmin's-striped frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Myobatrachidae;
OC   Limnodynastes.
RN   [1]
RP   SEQUENCE, AND MASS SPECTROMETRY.
RC   TISSUE=SKIN SECRETION;
RA   Bradford A.M., Raftery M.J., Bowie J.H., Wallace J.C., Tyler M.J.;
RT   "Peptides from Australian frogs. The structure of the dynastins from
RT   Limnodynastes salmini and Fletcherin from Limnodynastes fletcheri.";
RL   Aust. J. Chem. 46:1235-1244(1993).
CC   -!- MASS SPECTROMETRY: MW=786; METHOD=FAB.
KW   Amphibian skin.
SQ   SEQUENCE   8 AA;  786 MW;  7B58772455B05728 CRC64;


  Query Match             28.9%;  Score 13;  DB 13;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 ANLNL 9
           :|| :
Db       4 SNLGI 8


RESULT  17
Q9TWD6
ID   Q9TWD6       PRELIMINARY;      PRT;     9 AA.
AC   Q9TWD6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   LED-NPF-1=NEUROPEPTIDE F-RELATED PEPTIDE.
OS   Leptinotarsa decemlineata (Colorado potato beetle).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC   Cucujiformia; Chrysomelidae; Chrysomelinae; Leptinotarsa.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 96245438.
RA   Spittaels K., Verhaert P., Shaw C., Johnston R.N., Devreese B.,
RA   Van Beeumen J., De Loof A.;
RT   "Insect neuropeptide F (NPF)-related peptides: isolation from Colorado
RT   potato beetle (Leptinotarsa decemlineata) brain.";
RL   Insect Biochem. Mol. Biol. 26:375-382(1996).
SQ   SEQUENCE   9 AA;  1066 MW;  7E02340736C76864 CRC64;


  Query Match             28.9%;  Score 13;  DB 5;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 GANLNL 9
           |  | |
Db       3 GPQLRL 8


RESULT  18
Q63480
ID   Q63480       PRELIMINARY;      PRT;     7 AA.
AC   Q63480;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   TR4-NS ORPHAN RECEPTOR (FRAGMENT).
GN   TR4.
```

```
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96198747.
RA   Yoshikawa T., Makino S., Gao X.M., Xing G.Q., Chuang D.M.,
RA   Detera-Wadleigh S.D.;
RT   "Splice variants of rat TR4 orphan receptor: differential expression
RT   of novel sequences in the 5'-untranslated region and C-terminal
RT   domain.";
RL   Endocrinology 137:1562-1571(1996).
DR   EMBL; U59125; AAB02827.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   7 AA;  758 MW;  672AA87864005350 CRC64;


  Query Match             26.7%;  Score 12;  DB 11;  Length 7;
  Best Local Similarity   33.3%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        2 LSGANL 7
            : | :|
Db        2 IRGGDL 7


RESULT  19
Q86871
ID   Q86871      PRELIMINARY;      PRT;     7 AA.
AC   Q86871;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   VIRION STRUCTURAL PROTEIN (FRAGMENT).
GN   GENE III.
OS   Cauliflower mosaic virus.
OC   Viruses; Retroid viruses; Caulimovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=11/3-7;
RX   MEDLINE; 95053898.
RA   al-Kaff N., Covey S.N.;
RT   "Variation in biological properties of cauliflower mosaic virus
RT   clones.";
RL   J. Gen. Virol. 75:3137-3145(1994).
DR   EMBL; S75948; CAB33416.1; -.
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  744 MW;  672054444DC5B030 CRC64;


  Query Match             26.7%;  Score 12;  DB 12;  Length 7;
  Best Local Similarity   33.3%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        2 LSGANL 7
            :|  |:
Db        1 ISANNI 6


RESULT  20
Q15889
ID   Q15889      PRELIMINARY;      PRT;     8 AA.
AC   Q15889;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   (CLONE XP15H8B) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   Lee C.C., Yazdani A., Wehnert M., Bailey J., Couch L., Xiong M.,
RA   Coolbaugh M.I., Chinault C.A., Baldini A., Lindsay E.A., Zhao Z.Y.,
RA   Caskey C.T.H.;
RL   Hum. Mol. Genet. 0:0-0(0).
DR   EMBL; L32070; AAA73879.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  865 MW;  0474472325A761E7 CRC64;

  Query Match             26.7%;  Score 12;  DB 4;  Length 8;
  Best Local Similarity   42.9%;  Pred. No. 3e+05;
  Matches    3;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        2 LSGANLN 8
            |  : ||
Db        1 LHPSKLN 7


RESULT  21
Q15901
ID   Q15901      PRELIMINARY;      PRT;     8 AA.
AC   Q15901;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   (CLONE XP7B11B) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   Lee C.C., Yazdani A., Wehnert M., Bailey J., Couch L., Xiong M.,
RA   Coolbaugh M.I., Chinault C.A., Baldini A., Lindsay E.A., Zhao Z.Y.,
RA   Caskey C.T.H.;
RL   Hum. Mol. Genet. 0:0-0(0).
DR   EMBL; L32080; AAA73891.1; -.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  860 MW;  37D72878676729CB CRC64;


  Query Match             26.7%;  Score 12;  DB 4;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSG 4
            :| |
Db        2 FLPG 5


RESULT  22
Q9Y4J3
ID   Q9Y4J3      PRELIMINARY;      PRT;     8 AA.
AC   Q9Y4J3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   RUNT/68NT/MTG8 PROTEIN (FRAGMENT).
GN   RUNT/68NT/MTG8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE MARROW;
RX   MEDLINE; 95002916.
RA   Tighe J.E., Calabi F.;
RT   "Alternative, out-of-frame runt/MTG8 transcripts are encoded by the
RT   derivative (8) chromosome in the t(8;21) of acute myeloid leukemia
RT   M2.";
RL   Blood 84:2115-2121(1994).
```

```
DR   EMBL; S74094; AAD14973.2; -.
FT   NON_TER       1      1
SQ   SEQUENCE   8 AA;  929 MW;  30B764405B17244B CRC64;


  Query Match             26.7%;  Score 12;  DB 4;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          6 NLNL 9
              || :
Db          2 NLEI 5


RESULT 23
Q95213
ID   Q95213      PRELIMINARY;      PRT;     8 AA.
AC   Q95213;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   GERMLINE DH (DF) GENE (FRAGMENT).
CN   DF.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F-I/RGM;
RA   Mage R.G., Chen H.T., Alexander C.B., Chen F.F.;
RL   Mol. Immunol. 0:0-0(0).
DR   EMBL; U62585; AAB18735.1; -.
KW   MHC.
FT   NON_TER       1      1
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  845 MW;  5CA861B5AB58677B CRC64;


  Query Match             26.7%;  Score 12;  DB 7;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSG 4
              | :|
Db          4 YSTG 7


RESULT 24
Q66807
ID   Q66807      PRELIMINARY;      PRT;     8 AA.
AC   Q66807;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   5'UTR IN IRES SEQUENCES (ISOLATE TH222) (FRAGMENT).
OS   Echovirus 25.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Enterovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TH222;
RA   Bailly J.L., Borman A.M., Peigue-Lafeuille H., Kean K.M.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X90724; CAA62259.1; -.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  821 MW;  EFC1B5A2D6DDD876 CRC64;


  Query Match             26.7%;  Score 12;  DB 12;  Length 8;
  Best Local Similarity   40.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 GANLN 8
              || ::
Db          2 GAQVS 6


RESULT 25
Q9R7E8
ID   Q9R7E8      PRELIMINARY;      PRT;     9 AA.
AC   Q9R7E8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   KPSD PROTEIN (FRAGMENT).
GN   KPSD.
OS   Escherichia coli.
OG   Plasmid pCR3.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95180691.
RA   Rosenow C., Roberts I.S., Jann K.;
RT   "Isolation from recombinant Escherichia coli and characterization of
RT   CMP-Kdo synthetase, involved in the expression of the capsular K5
RT   polysaccharide (K-CKS).";
RL   FEMS Microbiol. Lett. 125:159-164(1995).
DR   EMBL; S76943; CAB33515.1; -.
KW   Plasmid.
FT   NON_TER       1      1
SQ   SEQUENCE   9 AA;  899 MW;  3EBBB72042C33DD8 CRC64;


  Query Match             26.7%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    3;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 GANLNL 9
              || : |
Db          2 GAKVIL 7


RESULT 26
Q9R635
ID   Q9R635      PRELIMINARY;      PRT;     9 AA.
AC   Q9R635;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   MAJOR OUTER MEMBRANE PROTEIN VARIABLE DOMAIN IV, MOMP VD IV.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92040090.
RA   Peterson E.M., Cheng X., Markoff B.A., Fielder T.J., de la Maza L.M.;
RT   "Functional and structural mapping of Chlamydia trachomatis species-
RT   specific major outer membrane protein epitopes by use of neutralizing
RT   monoclonal antibodies.";
RL   Infect. Immun. 59:4147-4153(1991).
SQ   SEQUENCE   9 AA;  976 MW;  9C61B041B7645361 CRC64;


  Query Match             26.7%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 LSG 4
              :||
Db          7 ISG 9


RESULT 27
Q14715
ID   Q14715      PRELIMINARY;      PRT;     9 AA.
```

```
AC   Q14715;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   KERATIN 14 (FRAGMENT).
GN   KRT14.
OS   Homo sapiens, (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92005680.
RA   Coulombe P., Hutton M., Letai A., Hebert A., Paller A., Fuchs E.;
RT   "Point mutations in human keratin 14 genes of epidermolysis bullosa
RT   simplex patients: genetic and functional analyses.";
RL   Cell 66:1301-1311(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95072587.
RA   Yamanishi K., Matsuki M., Konishi K., Yasuno H.;
RT   "A novel mutation of Leu122 to Phe at a highly conserved hydrophobic
RT   residue in the helix initiation motif of keratin 14 in epidermolysis
RT   bullosa simplex.";
RL   Hum. Mol. Genet. 3:1171-1172(1994).
DR   EMBL; D28807; BAA05967.1; -.
KW   Keratin.
FT   NON_TER       1      1
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1138 MW;  BE300AA449C456D6 CRC64;

  Query Match             26.7%;  Score 12;  DB 4;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       6 NLN 8
           | |
Db       4 NFN 6

RESULT  28
Q16220
ID   Q16220      PRELIMINARY;      PRT;     9 AA.
AC   Q16220;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HGRP PROTEIN (FRAGMENT).
GN   HGRP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94320083.
RA   Nagalla S.R., Spindel E.R.;
RT   "Functional analysis of the 5'-flanking region of the human gastrin-
RT   releasing peptide gene in small cell lung carcinoma cell lines.";
RL   Cancer Res. 54:4461-4467(1994).
DR   EMBL; S73265; AAD14116.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1070 MW;  77FE37672B040864 CRC64;

  Query Match             26.7%;  Score 12;  DB 4;  Length 9;
  Best Local Similarity   37.5%;  Pred. No. 3e+05;
  Matches    3;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 LSGANLNL 9
           : |  | |
Db       1 MRGRELPL 8
```

```
RESULT  29
Q9UCN5
ID   Q9UCN5      PRELIMINARY;      PRT;     9 AA.
AC   Q9UCN5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CELL-SURFACE HEPARIN/HEPARANSULFATE-BINDING PROTEIN PEPTIDE 2
DE   (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE.
RX   MEDLINE; 92291065.
RA   Raboudi N., Julian J., Rohde L.H., Carson D.D.;
RT   "Identification of cell-surface heparin/heparan sulfate-binding
RT   proteins of a human uterine epithelial cell line (RL95).";
RL   J. Biol. Chem. 267:11930-11939(1992).
SQ   SEQUENCE   9 AA;  1008 MW;  CB56D0544732C732 CRC64;

  Query Match             26.7%;  Score 12;  DB 4;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 LNL 9
           ||:
Db       5 LNI 7

RESULT  30
Q27396
ID   Q27396      PRELIMINARY;      PRT;     9 AA.
AC   Q27396;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   RHOPTRY ASSOCIATED PROTEIN 1.
GN   RAP-1.
OS   Babesia bovis.
OC   Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MO7;
RA   Suarez C.E., Palmer G.H., Hines S.A., McElwain T.F.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L77326; AAA96415.1; -.
SQ   SEQUENCE   9 AA;  931 MW;  6D5BD2C865B05044 CRC64;

  Query Match             26.7%;  Score 12;  DB 5;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       4 ISG 6

RESULT  31
Q28112
ID   Q28112      PRELIMINARY;      PRT;     9 AA.
AC   Q28112;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   6-PHOSPHOFRUCTO-2-KINASE/FRUCTOSE-2, 6-BISPHOSPHATASE (FRAGMENT).
GN   GENE B.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93387464.
RA   Vidal H., Crepin K.M., Rider M.H., Hue L., Rousseau G.G.;
RT   "Cloning and expression of novel isoforms of 6-phosphofructo-2-
RT   kinase/fructose-2,6-bisphosphatase from bovine heart.";
RL   FEBS Lett. 330:329-333(1993).
DR   EMBL; X74564; CAA52652.1; -.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  837 MW;  859CA5BDC7644865 CRC64;


  Query Match             26.7%;  Score 12;  DB 6;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       1 MSG 3


RESULT  32
Q89491
ID   Q89491      PRELIMINARY;      PRT;     9 AA.
AC   Q89491;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 1.1 KDA PROTEIN.
OS   Murine minute virus (Murine parvovirus).
OC   Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LYMPHOTROPIC VARIANT;
RX   MEDLINE; 86115415.
RA   Astell C.R., Gardiner E.M., Tattersall P.;
RT   "DNA sequence of the lymphotropic variant of minute virus of mice,
RT   MVM(i), and comparison with the DNA sequence of the fibrotropic
RT   prototype strain.";
RL   J. Virol. 570:656-669(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MVM(P);
RX   MEDLINE; 83143341.
RA   Astell C.R., Thomson M., Merchlinsky M., Ward D.C.;
RT   "The complete DNA sequence of minute virus of mice, an autonomous
RT   parvovirus.";
RL   Nucleic Acids Res. 11:999-1018(1983).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MVM(P);
RX   MEDLINE; 86115415.
RA   Astell C.R., Gardiner E.M., Tattersall P.;
RT   "DNA sequence of the lymphotropic variant of minute virus of mice,
RT   MVM(i), and comparison with the DNA sequence of the fibrotropic
RT   prototype strain.";
RL   J. Virol. 57:656-669(1986).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MVM(P);
RX   MEDLINE; 87061199.
RA   Morgan W.R., Ward D.C.;
RT   "Three splicing patterns are used to excise the small intron common to
RT   all minute virus of mice RNAs.";
RL   J. Virol. 60:1170-1174(1986).
DR   EMBL; M12032; AAA69570.1; -.
DR   EMBL; J02275; AAA67112.1; -.
DR   EMBL; V01115; CAA24311.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   9 AA;  1061 MW;  C3FD405863637862 CRC64;


  Query Match             26.7%;  Score 12;  DB 12;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GANL 7
           | |:
Db       6 GINV 9


RESULT  33
Q47505
ID   Q47505      PRELIMINARY;      PRT;     7 AA.
AC   Q47505;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   PLASMID PMCCC7 MCCA,B,C,D,E,F GENES.
GN   MCCA.
OS   Escherichia coli.
OG   Plasmid pMccC7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96099297.
RA   Gonzalez-Pastor J.E., San Millan J.L., Castilla M.A., Moreno F.;
RT   "Structure and organization of plasmid genes required to produce the
RT   translation inhibitor microcin C7.";
RL   J. Bacteriol. 177:7131-7140(1995).
DR   EMBL; X57583; CAA40808.1; -.
KW   Plasmid.
SQ   SEQUENCE   7 AA;  763 MW;  644DD44861B406F0 CRC64;


  Query Match             24.4%;  Score 11;  DB 2;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       6 NLN 8
           | |
Db       5 NAN 7


RESULT  34
Q07624
ID   Q07624      PRELIMINARY;      PRT;     7 AA.
AC   Q07624;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL PROTEIN (FRAGMENT).
OS   Rous sarcoma virus.
OC   Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PRAGUE C;
RX   MEDLINE; 93010967.
RA   Donze O., Spahr P.F.;
RT   "Role of the open reading frames of Rous sarcoma virus leader RNA in
RT   translation and genome packaging.";
RL   EMBO J. 11:3747-3757(1992).
DR   EMBL; X67587; CAA47862.1; -.
KW   Hypothetical protein.
FT   NON_TER       7      7
SQ   SEQUENCE   7 AA;  672 MW;  776045A7687DD6F0 CRC64;


  Query Match             24.4%;  Score 11;  DB 12;  Length 7;
  Best Local Similarity   16.7%;  Pred. No. 3e+05;
  Matches    1;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LSGANL 7
           ::| ::
Db       1 MAGPSI 6
```

```
RESULT  35
Q9YQ10
ID   Q9YQ10      PRELIMINARY;      PRT;     7 AA.
AC   Q9YQ10;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HYPOTHETICAL FUSION PROTEIN.
OS   porcine transmissible gastroenteritis virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99099045.
RA   Izeta A., Smerdou C., Alonso S., Penzes Z., Mendez A., Plana-Duran J.,
RA   Enjuanes L.;
RT   "Replication and packaging of transmissible gastroenteritis
RT   coronavirus-derived synthetic minigenomes.";
RL   J. Virol. 73:1535-1545(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 95159435.
RA   Eleouet J., Rasschaert D., Lambert P., Levy L., Vende P., Laude H.;
RT   "Complete sequence (20 kilobases) of the polyprotein-encoding gene 1
RT   of transmissible gastroenteritis virus.";
RL   Virology 206:817-822(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88078100.
RA   Rasschaert D., Gelfi J., Laude H.;
RT   "Enteric coronavirus TGEV: partial sequence of the genomic RNA, its
RT   organization and expression.";
RL   Biochimie 69:591-600(1987).
DR   EMBL; AJ011482; CAA09625.1; -.
SQ   SEQUENCE   7 AA;  927 MW;  69D6D7273B5726F0 CRC64;

  Query Match             24.4%;  Score 11;  DB 12;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 3e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 YL 2
          ||
Db      3 YL 4


Search completed: December 16, 2000, 04:22:12
Job time: 4606 sec
```

```
                    GenCore version 4.5
           Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 02:01:07 ; Search time 89.11 Seconds
                                              (without alignments)
                                              6.409 Million cell updates/sec

Title:          US-09-529-121-1
Perfect score:  45
Sequence:       1 YLSGANLNL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters:      793

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :       PIR_65:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                %
Result          Query
  No.   Score   Match Length DB  ID                   Description
-------------------------------------------------------------------------
    1     16    35.6      7  2  S58797               serine/threonine-s
    2     16    35.6      8  2  H41978               calliFMRFamide 8 -
    3     16    35.6      8  2  E60588               sperm-activating p
    4     15    33.3      8  2  T13818               cytochrome oxidase
    5     15    33.3      9  2  PT0288               Ig heavy chain CRD
    6     15    33.3      9  2  G41946               T-cell receptor ga
    7     14    31.1      6  2  PT0605               T-cell receptor be
    8     14    31.1      6  2  PT0593               T-cell receptor be
    9     14    31.1      7  2  PT0654               T-cell receptor be
   10     14    31.1      8  2  PN0043               phosphatidylethano
   11     14    31.1      9  2  A44873               caldesmon - rabbit
   12     14    31.1      9  2  C57444               neuropeptide Grb-A
   13     14    31.1      9  2  JQ0914               MHC class I histoc
   14     13    28.9      4  2  S43959               Ig mu chain V regi
   15     13    28.9      7  2  I50210               gene c-rel protein
   16     13    28.9      9  1  YFPG                 thymic factor - pi
   17     13    28.9      9  2  A61364               isotocin - common
   18     13    28.9      9  2  A60957               thymocyte growth p
   19     13    28.9      9  2  C41170               photosystem II pro
   20     12    26.7      5  2  S62883               seminal plasma pro
   21     12    26.7      6  2  I51434               H4 histone - Afric
   22     12    26.7      6  2  I49424               cytotoxic T-lympho
   23     12    26.7      7  1  XEYDGD               galactose oxidase
   24     12    26.7      7  2  PN0649               alpha-dextrin endo
   25     12    26.7      7  2  A34818               vicilin 72K chain
   26     12    26.7      8  2  PQ0701               unidentified 6.5/3
   27     12    26.7      8  2  PL0184               capsid protein VP-
   28     12    26.7      8  2  A21440               variant surface gl
   29     12    26.7      8  2  A41117               acetylcholinestera
   30     12    26.7      8  2  PT0725               T-cell receptor be
   31     12    26.7      9  2  S56004               glucan 1,3-beta-gl
   32     11    24.4      4  2  S43014               hypothetical prote
   33     11    24.4      5  1  HOROHA               proctolin - Americ
   34     11    24.4      5  2  A41225               copper resistance
   35     11    24.4      5  2  A60411               proctolin - Atlant
   36     11    24.4      5  2  C23751               spinal cord peptid
   37     11    24.4      6  2  JU0355               lipopeptide WS1279
   38     11    24.4      6  2  B44510               hypothetical prote
   39     11    24.4      6  2  PT0280               Ig heavy chain CRD
   40     11    24.4      7  2  S71867               glutathione transf
   41     11    24.4      7  2  PC2370               probable H+-transp
   42     11    24.4      7  2  S45311               microcin C7 - Esch
   43     11    24.4      7  2  S42407               gramicidin S synth
   44     11    24.4      7  2  I40504               hypothetical prote
   45     11    24.4      7  2  S78024               ribosomal protein
   46     11    24.4      7  2  H33098               180K exoantigen -
   47     11    24.4      7  2  C56793               platelet glycoprot
   48     11    24.4      7  2  PT0542               T-cell receptor be
   49     11    24.4      8  1  LFSAME               probable msrA lead
   50     11    24.4      8  2  G33098               205K exoantigen -
   51     11    24.4      8  2  JS0318               leucokinin VIII -
   52     11    24.4      8  2  A61467               penalbumin - Adeli
   53     11    24.4      8  2  PT0311               Ig heavy chain CRD
   54     11    24.4      8  2  PT0627               T-cell receptor be
   55     11    24.4      8  2  B47594               aspartate kinase (
   56     11    24.4      8  2  S68325               blood cell protein
   57     11    24.4      9  1  AKLQIM               locustamyoinhibiti
   58     11    24.4      9  2  A91466               oxytocin - hippopo
   59     11    24.4      9  2  A92774               oxytocin - spotted
   60     11    24.4      9  2  A93147               oxytocin - finback
   61     11    24.4      9  2  A93408               oxytocin - Austral
   62     11    24.4      9  2  B90667               oxytocin - rabbit
   63     11    24.4      9  2  A61230               calsequestrin, car
   64     11    24.4      9  2  S63491               dissimilatory sulf
   65     11    24.4      9  2  JP0073               ribosomal protein
   66     11    24.4      9  2  A43848               cell surface adhes
   67     11    24.4      9  2  A11497               transaldolase (EC
   68     11    24.4      9  2  A57444               neuropeptide Grb-A
   69     11    24.4      9  2  B57444               neuropeptide Grb-A
   70     11    24.4      9  2  JN0027               [Phe-6]-mosact - s
   71     11    24.4      9  2  I50633               c-rel protein - ch
   72     11    24.4      9  2  PT0268               Ig heavy chain CRD
   73     11    24.4      9  2  PH0935               T-cell receptor be
   74     11    24.4      9  2  PH0918               T-cell receptor be
   75     11    24.4      9  4  I57650               hemoglobin alpha c

                              ALIGNMENTS

RESULT   1
S58797
serine/threonine-specific protein kinase c-mos - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 23-Feb-1997
C;Accession: S58797
R;Nagao, Y.
Biochim. Biophys. Acta 1245, 130-143, 1995
A;Title: Expression of c-mos protein in cultured rat spermatogenic cells and evidence
A;Reference number: S58797; MUID:95383384
A;Accession: S58797
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-7 <NAG>
C;Genetics:
A;Gene: c-mos
C;Keywords: phosphotransferase; protein kinase


  Query Match             35.6%;  Score 16;  DB 2;  Length 7;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy       4 GANL 7
           | ||
Db       3 GGNL 6


RESULT   2
H41978
calliFMRFamide 8 - bluebottle fly (Calliphora vomitoria)
C;Species: Calliphora vomitoria
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C;Accession: H41978
R;Duve, H.; Johnsen, A.H.; Sewell, J.C.; Scott, A.G.; Orchard, I.; Rehfeld, J.F.; Thorpe
Proc. Natl. Acad. Sci. U.S.A. 89, 2326-2330, 1992
A;Title: Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH-2 neuropeptides (desi
A;Reference number: A41978; MUID:92196111
A;Accession: H41978
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-8 <DUV>
C;Keywords: amidated carboxyl end; neuropeptide
F;8/Modified site: amidated carboxyl end (Phe) #status experimental


  Query Match             35.6%;  Score 16;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GAN 6
           |||
Db       1 GAN 3


RESULT   3
E60588
sperm-activating peptide a - sea urchin (Pseudoboletia maculata)
N;Alternate names: speract homolog
C;Species: Pseudoboletia maculata
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 20-Jun-2000
C;Accession: E60588
R;Yoshino, K.I.; Kajiura, H.; Nomura, K.; Takao, T.; Shimonishi, Y.; Kurita, M.; Yamaguc
Comp. Biochem. Physiol. B 94, 739-751, 1989
A;Title: A halogenated amino acid-containing sperm activating peptide and its related pe
otus nudus, Echinometra mathaei and Heterocentrotus mammillatus.
A;Reference number: A60527
A;Accession: E60588
A;Molecule type: protein
A;Residues: 1-8 <YOS>


  Query Match             35.6%;  Score 16;  DB 2;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LSGAN 6
           | | |
Db       4 LDGVN 8


RESULT   4
T13818
cytochrome oxidase subunit I - Atlantic hagfish mitochondrion (fragment)
C;Species: mitochondrion Myxine glutinosa (Atlantic hagfish)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C;Accession: T13818
R;Delarbre, C.; Barriel, V.; Tillier, S.; Janvier, P.; Gachelin, G.
Mol. Biol. Evol. 14, 807-813, 1997
A;Title: The main features of the craniate mitochondrial DNA between the ND1 and the COI
A;Reference number: Z17775; MUID:97398704
A;Accession: T13818
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-8 <DEL>
```

```
A;Cross-references: EMBL:Y09527; NID:e1011465; PID:e329906; PIDN:CAA70718.1
C;Genetics:
A;Genome: mitochondrion
A;Note: COI
C;Keywords: mitochondrion


  Query Match             33.3%;  Score 15;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLS 3
           |||
Db       2 YLS 4


RESULT   5
PT0288
Ig heavy chain CRD3 region (clone 4-106) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C;Accession: PT0288
R;Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991
A;Title: Preferential utilization of specific immunoglobulin heavy chain diversity an
A;Reference number: PT0222; MUID:91108337
A;Accession: PT0288
A;Molecule type: DNA
A;Residues: 1-9 <YAM>
A;Experimental source: B lymphocyte
C;Keywords: heterotetramer; immunoglobulin


  Query Match             33.3%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           | ||
Db       5 YSSG 8


RESULT   6
G41946
T-cell receptor gamma chain (2t.23) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 07-May-1999
C;Accession: G41946
R;Whetsell, M.; Mosley, R.L.; Whetsell, L.; Schaefer, F.V.; Miller, K.S.; Klein, J.R.
Mol. Cell. Biol. 11, 5902-5909, 1991
A;Title: Rearrangement and junctional-site sequence analyses of T-cell receptor gamma
A;Reference number: A41946; MUID:92049316
A;Accession: G41946
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-9 <WHE>
C;Keywords: T-cell receptor


  Query Match             33.3%;  Score 15;  DB 2;  Length 9;
  Best Local Similarity   75.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           | ||
Db       5 YSSG 8


RESULT   7
PT0605
T-cell receptor beta chain V-D-J region (120-1L) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
```

```
C;Accession: PT0605
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0605
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-6 <FEE>
A;Experimental source: newborn thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGA 5
           |||
Db       2 SGA 4


RESULT   8
PT0593
T-cell receptor beta chain V-D-J region (159-1F) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0593
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0593
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-6 <FEE>
A;Experimental source: day 19 fetal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGA 5
           |||
Db       4 SGA 6


RESULT   9
PT0654
T-cell receptor beta chain V-D-J region (121-1BK) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0654
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0654
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-7 <FEE>
A;Experimental source: day 4 postnatal thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             31.1%;  Score 14;  DB 2;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGA 5
           |||
```

```
Db       2 SGA 4


RESULT  10
PN0043
phosphatidylethanol amine-binding protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 29-Oct-1997 #sequence_revision 29-Oct-1997 #text_change 23-Jan-1998
C;Accession: PN0043
R;Kato, H.
Kawasaki Igakkaishi 22, 245-259, 1996
A;Title: Analysis of proteins isolated by two dimensional electrophoresis of mouse ne
A;Reference number: PN0041
A;Accession: PN0043
A;Molecule type: protein
A;Residues: 1-8 <KAT>
A;Experimental source: neuroblastoma cell
C;Comment: The molecular mass is 23,500 and the pI is 4.8. The amino-terminus is bloc
C;Keywords: brain


  Query Match             31.1%;  Score 14;  DB 2;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           |||
Db       5 LSG 7


RESULT  11
A44873
caldesmon - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 31-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 31-Oct-1997
C;Accession: A44873
R;Ikebe, M.; Hornick, T.
Arch. Biochem. Biophys. 288, 538-542, 1991
A;Title: Determination of the phosphorylation sites of smooth muscle caldesmon by pro
A;Reference number: A44873; MUID:91378498
A;Accession: A44873
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <IKE>
A;Experimental source: skeletal myosin
A;Note: sequence extracted from NCBI backbone (NCBIP:63199)
C;Superfamily: caldesmon


  Query Match             31.1%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   33.3%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GANLNL 9
           |::| :
Db       1 GSSLKI 6


RESULT  12
C57444
neuropeptide Grb-AST B3 - two-spotted cricket
C;Species: Gryllus bimaculatus (two-spotted cricket)
C;Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 26-Jan-1996
C;Accession: C57444
R;Lorenz, M.W.; Kellner, R.; Hoffmann, K.H.
J. Biol. Chem. 270, 21103-21108, 1995
A;Title: A family of neuropeptides that inhibit juvenile hormone biosynthesis in the
A;Reference number: A57444; MUID:95403341
A;Accession: C57444
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <LOR>
```

```
  Query Match             31.1%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           |||
Db       5 LSG 7


RESULT  13
JQ0914
MHC class I histocompatibility antigen heavy chain - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 24-Feb-1995
C;Accession: JQ0914
R;Kanki, T.; Kuwasawa, N.; Sekiya, Y.; Ichikawa, Y.
submitted to JIPID, May 1991
A;Description: Responsive expression of a MHC class I epitope and genes following Marek'
A;Reference number: JQ0914
A;Accession: JQ0914
A;Molecule type: mRNA
A;Residues: 1-9 <KAN>
A;Experimental source: kidney, strain cornell N


  Query Match             31.1%;  Score 14;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGAN 6
           :|:|
Db       3 TGSN 6


RESULT  14
S43959
Ig mu chain V region (clone 13) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 20-Oct-1994 #sequence_revision 26-Jul-1996 #text_change 17-Mar-1999
C;Accession: S43959
R;Wagner, S.D.; Williams, G.T.; Larson, T.; Neuberger, M.S.; Kitamura, D.; Rajewsky, K.;
Nucleic Acids Res. 22, 1389-1393, 1994
A;Title: Antibodies generated from human immunoglobulin miniloci in transgenic mice.
A;Reference number: S43956; MUID:94248036
A;Accession: S43959
A;Molecule type: DNA
A;Residues: 1-4 <WAG>
C;Keywords: immunoglobulin


  Query Match             28.9%;  Score 13;  DB 2;  Length 4;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSG 4
           | :|
Db       1 YCAG 4


RESULT  15
I50210
gene c-rel protein - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 28-Feb-1997
C;Accession: I50210
R;Kabrun, N.; Bumstead, N.; Hayman, M.J.; Enrietto, P.J.
Mol. Cell. Biol. 10, 4788-4794, 1990
A;Title: Characterization of a novel promoter insertion in the c-rel locus.
A;Reference number: I50210; MUID:90355995
A;Accession: I50210
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
```

```
A;Residues: 1-7 <KAB>
A;Cross-references: GB:M55577; NID:g555438; PID:g211661
C;Genetics:
A;Gene: c-rel


  Query Match             28.9%;  Score 13;  DB 2;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGA 5
           ::||
Db       1 MAGA 4


RESULT  16
YFPG
thymic factor - pig
N;Alternate names: FTS (facteur thymique serique)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 07-May-1999
C;Accession: A01523; A60983
R;Pleau, J.M.; Dardenne, M.; Blouquit, Y.; Bach, J.F.
J. Biol. Chem. 252, 8045-8047, 1977
A;Title: Structural study of circulating thymic factor: a peptide isolated from pig s
A;Reference number: A01523; MUID:78026571
A;Accession: A01523
A;Molecule type: protein
A;Residues: 1-9 <PLE>
R;Bach, J.F.; Dardenne, M.; Pleau, J.M.; Rosa, J.
Nature 266, 55-57, 1977
A;Title: Biochemical characterisation of a serum thymic factor.
A;Reference number: A60983; MUID:77123829
A;Accession: A60983
A;Molecule type: protein
A;Residues: 'Z',2-4,'Z',6-9 <BAC>
C;Comment: This peptide induces DNA synthesis in immature thymocytes, but not periphe
in a variety of immunoassays.
C;Comment: See PIR:A60957 (sheep) for discussion of another possible N-terminal modif
C;Superfamily: thymic factor
C;Keywords: pyroglutamic acid
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental


  Query Match             28.9%;  Score 13;  DB 1;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GAN 6
           |:|
Db       7 GSN 9


RESULT  17
A61364
isotocin - common carp
C;Species: Cyprinus carpio (common carp)
C;Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 20-Jun-2000
C;Accession: A61364
R;Acher, R.; Chauvet, J.; Chauvet, M.T.; Crepy, D.
Comp. Biochem. Physiol. A 14, 245-254, 1965
A;Title: Caracterisation des hormones neurohypophysaires d'un poisson osseux d'eau do
A;Reference number: A61364
A;Accession: A61364
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <ACH>
C;Superfamily: oxytocin-neurophysin
C;Keywords: amidated carboxyl end; neuropeptide; posterior pituitary
F;9/Modified site: amidated carboxyl end (Gly) #status experimental


  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
```

```
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLS 3
           |:|
Db       2 YIS 4


RESULT  18
A60957
thymocyte growth peptide - sheep
N;Contains: FTS (facteur thymique serique)
C;Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 24-Nov-1999
C;Accession: A60957
R;Ernstroem, U.; Gafvelin, G.; Rudja, J.M.
Biosci. Rep. 10, 403-412, 1990
A;Title: Purification of thymocyte growth peptide (TGP) from sheep thymus. Relationship
A;Reference number: A60957; MUID:91064427
A;Accession: A60957
A;Molecule type: protein
A;Residues: 1-9 <ERN>
C;Comment: This peptide induces DNA synthesis in immature thymocytes, but not peripheral
in a variety of immunoassays.
C;Comment: This peptide was isolated in two forms. One form contained the pyrrolidone ca
r form (thymocyte growth peptide) contains a large, non-peptide blocking group with a hi
C;Superfamily: thymic factor
C;Keywords: blocked amino end; pyroglutamic acid
F;1/Modified site: pyrrolidone carboxylic acid (Glx) (in FTS) #status experimental
F;1/Modified site: blocked amino end (Glx) (in thymocyte growth peptide) #status experim


  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GAN 6
           |:|
Db       7 GSN 9


RESULT  19
C41170
photosystem II protein psbM - Chlamydomonas reinhardtii (fragment)
C;Species: Chlamydomonas reinhardtii
C;Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 30-Sep-1993
C;Accession: C41170
R;de Vitry, C.; Diner, B.A.; Popot, J.L.
J. Biol. Chem. 266, 16614-16621, 1991
A;Title: Photosystem II particles from Chlamydomonas reinhardtii. Purification, molecula
A;Reference number: A41170; MUID:91358452
A;Accession: C41170
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-9 <DE5>


  Query Match             28.9%;  Score 13;  DB 2;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGA 5
           ::||
Db       4 IAGA 7


RESULT  20
S62883
seminal plasma protein II - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C;Accession: S62883
R;Romero, A.; Varela, P.F.; Sanz, L.; Toepfer-Petersen, E.; Calvete, J.J.
```

```
FEBS Lett. 382, 15-17, 1996
A;Title: Crystallization and preliminary X-ray diffraction analysis of boar seminal p
A;Reference number: S62882; MUID:96196555
A;Accession: S62883
A;Molecule type: protein
A;Residues: 1-5 <ROM>
C;Complex: heterodimer; seminal plasma protein I and seminal plasma protein II
C;Keywords: glycoprotein; heterodimer; semen


  Query Match             26.7%;  Score 12;  DB 2;  Length 5;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 ANLN 8
           | :|
Db       1 ARIN 4


RESULT  21
I51434
H4 histone - African clawed frog (fragment)
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Sep-1996
C;Accession: I51434
R;Woodland, H.R.; Warmington, J.R.; Ballantine, J.E.M.; Turner, P.C.
Nucleic Acids Res. 12, 4939-4958, 1984
A;Title: Are there major developmentally regulated H4 gene classes in Xenopus?.
A;Reference number: I51391; MUID:84247348
A;Accession: I51434
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-6 <WOO>
A;Cross-references: GB:K02304; NID:g214227; PID:g555517


  Query Match             26.7%;  Score 12;  DB 2;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       1 MSG 3


RESULT  22
I49424
cytotoxic T-lymphocyte proteinase 3 (EC 3.4.21.-) - western wild mouse (fragment)
C;Species: Mus spretus (western wild mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C;Accession: I49424
R;Ko, M.S.; Wang, X.; Horton, J.H.; Hagen, M.D.; Takahashi, N.; Maezaki, Y.; Nadeau,
Mamm. Genome 5, 349-355, 1994
A;Title: Genetic mapping of 40 cDNA clones on the mouse genome by PCR.
A;Reference number: I48934; MUID:94319082
A;Accession: I49424
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6 <RES>
A;Cross-references: EMBL:U05745; NID:g497084; PIDN:AAB60481.1; PID:g642831
C;Keywords: hydrolase; serine proteinase


  Query Match             26.7%;  Score 12;  DB 2;  Length 6;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       6 NLNL 9
           |: |
Db       2 NMKL 5


RESULT  23
```

```
XEYDGD
galactose oxidase inhibitor - fungus (Cladobotryum dendroides)
C;Species: Cladobotryum dendroides
C;Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 31-Dec-1993
C;Accession: A01341
R;Avigad, G.; Markus, Z.
Fed. Proc. 31, 447, 1972
A;Reference number: A01341
A;Accession: A01341
A;Molecule type: protein
A;Residues: 1-7 <AVI>
C;Comment: The mycelia of this imperfect fungus produce the metalloenzyme galactose oxid
apoenzyme, may inactivate the enzyme by binding to its prosthetic copper group.
C;Superfamily: galactose oxidase inhibitor
C;Keywords: copper


  Query Match             26.7%;  Score 12;  DB 1;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        3 SGAN 6
            :| |
Db        1 AGQN 4

RESULT  24
PN0649
alpha-dextrin endo-1,6-alpha-glucosidase (EC 3.2.1.41) - Bacillus sp. (strain S-1) (frag
C;Species: Bacillus sp.
C;Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 07-May-1999
C;Accession: PN0649
R;Kim, C.H.; Choi, H.I.; Lee, D.S.
Biosci. Biotechnol. Biochem. 57, 1632-1637, 1993
A;Title: Purification and biochemical properties of an alkaline pullulanase from alkalop
A;Reference number: PN0649; MUID:94080025
A;Accession: PN0649
A;Molecule type: protein
A;Residues: 1-7 <KIM>
C;Comment: This enzyme is used togther with glucoamylase to improve the efficiency of st
ntent in high maltose syrups.
C;Keywords: glycosidase; hydrolase; polysaccharide degradation


  Query Match             26.7%;  Score 12;  DB 2;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        7 LNL 9
            ||:
Db        2 LNM 4

RESULT  25
A34818
vicilin 72K chain - pigeon pea (fragment)
C;Species: Cajanus cajan (pigeon pea)
C;Date: 13-Jul-1990 #sequence_revision 13-Jul-1990 #text_change 30-Sep-1993
C;Accession: A34818
R;Mawal, Y.R.; Mawal, M.R.; Ranjekar, P.K.
Biochem. Biophys. Res. Commun. 166, 1446-1452, 1990
A;Title: Unusual denaturation properties of vicilin from Cajanus cajan.
A;Reference number: A34818; MUID:90165956
A;Accession: A34818
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-7 <MAW>


  Query Match             26.7%;  Score 12;  DB 2;  Length 7;
  Best Local Similarity   40.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy        4 GANLN 8
            || ::
Db        1 GARVD 5

RESULT  26
PQ0701
unidentified 6.5/31K protein [imported] - rice (fragment)
C;Species: Oryza sativa (rice)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: PQ0701
R;Komatsu, S.; Kajiwara, H.; Hirano, H.
Theor. Appl. Genet. 86, 935-942, 1993
A;Title: A rice protein library; a data-file of rice proteins separated by two-dimens
A;Reference number: PQ0696
A;Accession: PQ0701
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-8 <KOM>


  Query Match             26.7%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        2 LSGA 5
            ::||
Db        5 VTGA 8

RESULT  27
PL0184
capsid protein VP-1 - murine poliovirus (fragment)
C;Species: murine poliovirus, Theiler's encephalomyelitis virus
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Feb-1995
C;Accession: PL0184
R;Zurbriggen, A.; Hogle, J.M.; Fujinami, R.S.
J. Exp. Med. 170, 2037-2049, 1989
A;Title: Alteration of amino acid 101 within capsid protein VP-1 changes the pathogen
A;Reference number: PL0184; MUID:90063468
A;Accession: PL0184
A;Molecule type: genomic RNA
A;Residues: 1-8 <ZUR>
C;Keywords: capsid protein


  Query Match             26.7%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        3 SGANLN 8
            ||   |
Db        1 SGGITN 6

RESULT  28
A21440
variant surface glycoprotein pSLc1 - Trypanosoma brucei (fragment)
C;Species: Trypanosoma brucei
C;Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 20-Mar-1998
C;Accession: A21440
R;Parsons, M.; Nelson, R.G.; Watkins, K.P.; Agabian, N.
Cell 38, 309-316, 1984
A;Title: Trypanosome mRNAs share a common 5' spliced leader sequence.
A;Reference number: A90853; MUID:84282716
A;Accession: A21440
A;Molecule type: mRNA
A;Residues: 1-8 <PAR>
A;Cross-references: GB:K02195; NID:g162150; PID:g162151
C;Keywords: glycoprotein


  Query Match             26.7%;  Score 12;  DB 2;  Length 8;
```

```
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSG 4
           :||
Db       1 MSG 3


RESULT  29
A41117
acetylcholinesterase (EC 3.1.1.7), venom - Asian cobra (fragment)
C;Species: Naja naja oxiana (Asian cobra, Oxus cobra)
C;Date: 27-Mar-1992 #sequence_revision 27-Mar-1992 #text_change 23-Jun-1993
C;Accession: A41117
R;Kreienkamp, H.J.; Weise, C.; Raba, R.; Aaviksaar, A.; Hucho, F.
Proc. Natl. Acad. Sci. U.S.A. 88, 6117-6121, 1991
A;Title: Anionic subsites of the catalytic center of acetylcholinesterase from Torpedo a
A;Reference number: A41117; MUID:91296772
A;Accession: A41117
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-8 <KRE>
C;Keywords: carboxylic ester hydrolase


  Query Match             26.7%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 GANL 7
           || :
Db       1 GAEM 4


RESULT  30
PT0725
T-cell receptor beta chain V-D-J region (140-2I) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C;Accession: PT0725
R;Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.
A;Reference number: PT0509; MUID:91277601
A;Accession: PT0725
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-8 <FEE>
A;Experimental source: newborn thymus, strain BALB/c
C;Keywords: T-cell receptor


  Query Match             26.7%;  Score 12;  DB 2;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 1.8e+05;
  Matches    3;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 SGANL 7
           ||  |
Db       2 SGDGL 6


RESULT  31
S56004
glucan 1,3-beta-glucosidase (EC 3.2.1.58) GNIII, extracellular - fungus (Acremonium pers
N;Alternate names: (1-3)-beta-D-glucan glucohydrolase GNIII
C;Species: Acremonium persicinum
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 07-May-1999
C;Accession: S56004
R;Pitson, S.M.; Seviour, R.J.; McDougall, B.M.; Woodward, J.R.; Stone, B.A.
Biochem. J. 308, 733-741, 1995
A;Title: Purification and characterization of three extracellular (1->3)-beta-D-glucan g
A;Reference number: S56002; MUID:97104268
A;Accession: S56004
```

```
A;Molecule type: protein
A;Residues: 1-9 <PIT>
C;Keywords: glycosidase; hydrolase


  Query Match             26.7%;  Score 12;  DB 2;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ANL 7
           ||:
Db       7 ANI 9


RESULT  32
S43014
hypothetical protein URF-2Y - Yersinia enterocolitica transposon TN3926
C;Species: Yersinia enterocolitica
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
C;Accession: S43014
R;Osbourn, S.E.V.; Turner, A.K.; Grinsted, J.
submitted to the EMBL Data Library, March 1994
A;Description: The structure of the bacterial transposable element, Tn3926.
A;Reference number: S43011
A;Accession: S43014
A;Molecule type: DNA
A;Residues: 1-4 <OSB>
A;Cross-references: EMBL:X78059; NID:g460067; PIDN:CAA54980.1; PID:g581836
C;Genetics:
A;Mobile element: transposon TN3926


  Query Match             24.4%;  Score 11;  DB 2;  Length 4;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       6 NLN 8
           | |
Db       2 NAN 4


RESULT  33
HOROHA
proctolin - American cockroach
C;Species: Periplaneta americana (American cockroach)
C;Date: 29-Jul-1981 #sequence_revision 29-Jul-1981 #text_change 23-Aug-1996
C;Accession: A01644
R;Starratt, A.N.; Brown, B.E.
Life Sci. 17, 1253-1256, 1975
A;Title: Structure of the pentapeptide proctolin, a proposed neurotransmitter in inse
A;Reference number: A93048; MUID:76074708
A;Accession: A01644
A;Molecule type: protein
A;Residues: 1-5 <STA>
A;Note: the synthetic peptide had the same chromatographic, electrophoretic, and phar
R;O'Shea, M.; Adams, M.E.
Science 213, 567-569, 1981
A;Title: Pentapeptide (proctolin) associated with an identified neuron.
A;Reference number: A94260; MUID:81225865
A;Contents: annotation; biological source
C;Comment: This peptide is found in the lateral white neurons, which occur (in the co
innervate the striated hindgut muscles in insects and stimulate contraction of these
C;Superfamily: proctolin
C;Keywords: neuropeptide


  Query Match             24.4%;  Score 11;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YL 2
           ||
Db       2 YL 3
```

```
RESULT  34
A41225
copper resistance protein A - Pseudomonas syringae pv. tomato (fragment)
C;Species: Pseudomonas syringae pv. tomato
C;Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 24-Jun-1993
C;Accession: A41225
R;Cha, J.S.; Cooksey, D.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 8915-8919, 1991
A;Title: Copper resistance in Pseudomonas syringae mediated by periplasmic and outer mem
A;Reference number: A41225; MUID:92020961
A;Accession: A41225
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-5 <CHA>


  Query Match             24.4%;  Score 11;  DB 2;  Length 5;
  Best Local Similarity   66.7%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGA 5
           ||:
Db       2 SGS 4


RESULT  35
A60411
proctolin - Atlantic horseshoe crab
C;Species: Limulus polyphemus (Atlantic horseshoe crab)
C;Date: 03-Feb-1993 #sequence_revision 03-Feb-1993 #text_change 17-Mar-1999
C;Accession: A60411
R;Groome, J.R.; Tillinghast, E.K.; Townley, M.A.; Vetrovs, A.; Watson III, W.H.; Hunt, D
Peptides 11, 205-211, 1990
A;Title: Identification of proctolin in the central nervous system of the horseshoe crab
A;Reference number: A60411; MUID:90287800
A;Accession: A60411
A;Molecule type: protein
A;Residues: 1-5 <GRO>
C;Comment: This neuropeptide stimulates cardiac output and hindgut motility in the horse
C;Keywords: neuropeptide


  Query Match             24.4%;  Score 11;  DB 2;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YL 2
           ||
Db       2 YL 3


Search completed: December 16, 2000, 03:35:12
Job time: 5645 sec
```

```
                         GenCore version 4.5
              Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 16, 2000, 00:52:59 ; Search time 107.12 Seconds
                                             (without alignments)
                                             2.873 Million cell updates/sec

Title:          US-09-529-121-1
Perfect score:  45
Sequence:       1 YLSGANLNL 9

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters:      61695

Minimum DB seq length: 0
Maximum DB seq length: 9

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 75 summaries

Database :      A_Geneseq_36:*
               1:  /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
               2:  /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
               3:  /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
               4:  /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
               5:  /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
               6:  /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
               7:  /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
               8:  /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
               9:  /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
              10:  /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
              11:  /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
              12:  /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
              13:  /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
              14:  /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
              15:  /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
              16:  /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
              17:  /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
              18:  /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
              19:  /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
              20:  /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
              21:  /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.

                              SUMMARIES
                 %
Result          Query
  No.   Score   Match Length DB  ID                  Description
-----------------------------------------------------------------------
    1      45   100.0      9  18  W39723              Human carcina-embr
    2      45   100.0      9  19  W77134              CEA synthetic pept
    3      45   100.0      9  19  W70045              CEA derived HLA-A2
    4      45   100.0      9  20  Y47655              Immunogenic peptid
    5      45   100.0      9  20  Y09525              Carcinoembryonic a
    6      43    95.6      9  20  Y09528              Carcinoembryonic a
    7      42    93.3      9  21  Y54173              HLA binding peptid
    8      40    88.9      9  20  Y09526              Carcinoembryonic a
    9      38    84.4      9  17  W00680              Peptide comprising
   10      38    84.4      9  20  Y09527              Carcinoembryonic a
   11      36    80.0      9  20  Y09529              Carcinoembryonic a
   12      29    64.4      9  17  W00690              NCA analogue of re
   13      22    48.9      9  19  W40267              K. oxytoca R-speci
   14      21    46.7      6  11  R09414              LFA-1 alpha subuni
   15      21    46.7      6  17  R88476              Internal tryptic p
   16      21    46.7      6  20  Y23411              V beta 6 clone fou
   17      21    46.7      8  13  R29031              Peptide encoded by
   18      21    46.7      8  16  R78208              B. thuringiensis e
   19      21    46.7      9  14  R37221              IL-6 antagonist pe
   20      21    46.7      9  15  R73808              Antigen fragment 1
   21      21    46.7      9  19  W54298              Human cytohesin-1
   22      20    44.4      7  20  Y41847              Rheumatoid arthrit
   23      20    44.4      8  14  R34629              B. thuringiensis d
   24      20    44.4      8  16  R78218              B. thuringiensis e
   25      20    44.4      8  18  W36092              E. coli DNA polyme
   26      20    44.4      9  16  R69993              Nonameric mimotope
   27      20    44.4      9  18  W38383              Synthetic pMEL17 p
   28      20    44.4      9  18  W19861              Fragment of enzyme
   29      20    44.4      9  20  Y55448              HLA binding plu-1
   30      20    44.4      9  20  Y55529              HLA binding plu-1
   31      20    44.4      9  20  Y55623              HLA binding plu-1
   32      20    44.4      9  20  Y46533              Immunogenic peptid
   33      20    44.4      9  20  Y47062              Immunogenic peptid
   34      20    44.4      9  20  Y47532              Immunogenic peptid
   35      19    42.2      7  15  R46809              Phytase derived pe
   36      19    42.2      8  15  R46700              Monomeric peptide
   37      19    42.2      8  16  R74536              Protease Nexin-1 P
   38      19    42.2      8  19  W57004              Enzyme inhibitor p
   39      19    42.2      9  14  R41929              Peptide fragment o
   40      19    42.2      9  16  R87430              Human MHC class II
   41      19    42.2      9  17  W49339              Human leucocyte an
   42      19    42.2      9  17  W49340              Human leucocyte an
   43      19    42.2      9  19  W70078              B. stearothermophi
   44      19    42.2      9  20  W96297              VatC peptide fragm
   45      19    42.2      9  21  Y56575              Tyrosinase 2 deriv
   46      18    40.0      5  20  Y28182              Factor Xa inhibiti
   47      18    40.0      6  21  Y83831              Pyrrolobenzodiazep
   48      18    40.0      7  19  Y21209              Human bcl2 proto-o
   49      18    40.0      7  20  Y41846              Rheumatoid arthrit
   50      18    40.0      7  20  Y17023              Heat shock protein
   51      18    40.0      7  21  Y54382              Amino acid sequenc
   52      18    40.0      8  20  Y16852              Heat shock protein
   53      18    40.0      8  20  W94102              VIF-derived HIV pr
   54      18    40.0      8  20  W82429              PVX-2 viral replic
   55      18    40.0      9  11  R07966              Tryptic fragment T
   56      18    40.0      9  13  R22425              Peptide derived fr
   57      18    40.0      9  15  R59233              Peptide fragment (
   58      18    40.0      9  15  R73856              Antigen fragment 1
   59      18    40.0      9  15  R73809              Antigen fragment 1
   60      18    40.0      9  16  R87433              Human MHC class II
   61      18    40.0      9  16  R70067              Control peptide 92
   62      18    40.0      9  17  R87006              Class II MHC antig
   63      18    40.0      9  18  W38393              Synthetic pMEL17 p
   64      18    40.0      9  18  W15173              Chlamydia trachoma
   65      18    40.0      9  18  W15174              Chlamydia trachoma
   66      18    40.0      9  19  W54299              Human cytohesin-1
   67      18    40.0      9  19  W54515              Synthetic polypept
   68      18    40.0      9  20  Y55366              HLA binding plu-1
   69      18    40.0      9  20  Y55506              HLA binding plu-1
   70      18    40.0      9  20  Y55507              HLA binding plu-1
   71      18    40.0      9  20  Y55610              HLA binding plu-1
   72      18    40.0      9  20  Y41907              Rheumatoid arthrit
   73      18    40.0      9  20  Y42029              Rheumatoid arthrit
   74      18    40.0      9  20  Y42038              Rheumatoid arthrit
   75      18    40.0      9  20  Y31774              Mycobacterial hsp6

                              ALIGNMENTS

RESULT   1
W39723
ID   W39723 standard; peptide; 9 AA.
XX
AC   W39723;
```

```
XX
DT   11-JUN-1998  (first entry)
XX
DE   Human carcina-embryonic antigen (CEA) peptide (pos. 571-579).
XX
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
XX
OS   Homo sapiens.
XX
PN   WO9741440-A1.
XX
PD   06-NOV-1997.
XX
PF   28-APR-1997;   97WO-NL00229.
XX
PR   23-DEC-1996;   96EP-0203670.
PR   26-APR-1996;   96EP-0201145.
XX
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
XX
PI   Kast WM,  Melief CJM,  Offringa R,  Toes REM,  Van Der Burg SH;
XX
DR   WPI; 1997-549891/50.
XX
PT   Method of selecting T cell peptide epitope(s) - by measuring the
PT   stability of HLA class I-peptide complexes on intact B cells
XX
PS   Example 3; Page 85; 109pp; English.
XX
CC   Peptides W39430-W39734 are used in a novel method for the selection of
CC   immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC   method involves the identification of peptide sequences capable of
CC   binding to an HLA (human leukocyte antigen) class I molecule and
CC   measuring the binding of this epitope peptide to the HLA class I peptide.
CC   The stability of binding of the peptide and MHC (major histocompatablilty
CC   complex) class I molecule is measured on intact human B cells carrying
CC   the MHC molecule at their cell surfaces. The method can be used to select
CC   peptide epitopes for generating vaccines against a disease associated
CC   with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC   especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC   immune responses. Peptide W39723 is derived from the human
CC   carcino-embryonic antigen (CEA) and has the ability to bind to the human
CC   MHC Class I allele HLA-A2.1.
XX
SQ   Sequence   9 AA;

  Query Match             100.0%;  Score 45;  DB 18;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGANLNL 9
            |||||||||
Db        1 ylsganlnl 9

RESULT   2
W77134
ID   W77134 standard; peptide; 9 AA.
XX
AC   W77134;
XX
DT   16-NOV-1998  (first entry)
XX
DE   CEA synthetic peptide epitope 1.
XX
KW   Tyrosinase; tyrosinase cytotoxic lymphocyte response;
KW   cytotoxic T lymphocyte; cysteine-depleted; melanoma.
XX
OS   Synthetic.
```

```
XX
PN   WO9833810-A2.
XX
PD   06-AUG-1998.
XX
PF   29-JAN-1998;   98WO-US01592.
XX
PR   30-JAN-1997;   97US-0037781.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI   Engelhard VH,  Hunt DF,  Kittlesen D,  Slingluff CL;
XX
DR   WPI; 1998-437388/37.
XX
PT   Disease specific immunogen  - comprises disease specific cytotoxic T
PT   lymphocyte epitope used to elicit melanoma specific CTL response
XX
PS   Disclosure; Page 27; 93pp; English.
XX
CC   The peptide epitope W77119-W77138 were created for human tumour-specific
CC   cytotoxic T lymphocyte response.  These peptides are are cysteine-
CC   depleted mutants of a native disease-specific CTL epitope.  The cysteine-
CC   depleted CTL epitopes elicit a stronger or more specific CTL response
CC   than the native epitope.  The epitopes can be used in a disease-specific
CC   immunogen to protect a mammal against disease in particular melanomas.
CC   The peptides may also be used to screen a sample for the presence of
CC   an antigen with the same epitope, or with a different cross-reactive
CC   epitope.
XX
SQ   Sequence   9 AA;

  Query Match             100.0%;  Score 45;  DB 19;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGANLNL 9
            |||||||||
Db        1 ylsganlnl 9

RESULT   3
W70045
ID   W70045 standard; peptide; 9 AA.
XX
AC   W70045;
XX
DT   22-OCT-1998  (first entry)
XX
DE   CEA derived HLA-A2.1 binding peptide 2 (residues 605-613).
XX
KW   Cytotoxic T lymphocyte; CTL; major histocompatibility complex; MHC;
KW   human leukocyte antigen; HLA; tumour associated antigen; cancer;
KW   antigen presenting cell; APC; immunogenic peptide; immune disorder;
KW   viral infection; AIDS; hepatitis; bacterial infection; malaria; CEA;
KW   fungal infection; tuberculosis; melanoma; carcinoembryonic antigen.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9833888-A1.
XX
PD   06-AUG-1998.
XX
PF   30-JAN-1998;   98WO-US01959.
XX
PR   31-JAN-1997;   97US-0036696.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Celis E,  Sette A,  Sidney J,  Southwood S,  Tsai V;
XX
```

```
DR   WPI; 1998-437445/37.
XX
PT   Production of antigen-specific cytotoxic T cells - by incubating
PT   immunogenic peptide(s) from antigen that binds class I major
PT   histocompatibility complex molecules with pre-treated antigen
PT   presenting cells
XX
PS   Example 6; Page 75; 104pp; English.
XX
CC   Sequences shown in W70044 to W70052 represent peptides derived from
CC   carcinoembryonic antigen (CEA). The peptides can bind to a human
CC   leukocyte antigen (HLA), HLA-A2.1 and are used to exemplify the method
CC   of invention of producing antigen-specific cytotoxic T cells (CTLs) in
CC   vitro. The method comprises contacting immunogenic peptides from an
CC   antigen that binds class I major histocompatibility complex (MHC)
CC   molecules with antigen presenting cells (APCs) pretreated with
CC   pretreatment growth factors, and incubating the APCs with purified CD8
CC   cells in the presence of at least 2 incubation growth factors, thereby
CC   producing antigen-specific CTLs. A method for specifically killing
CC   target cells in a human patient is also provided which comprises
CC   obtaining a fluid sample containing CTLs from a patient, contacting the
CC   cytotoxic T cells with APCs pretreated with pre-treatment growth factors,
CC   where the APCs comprise class I MHC molecules. The pretreated APCs are
CC   incubated with the cytotoxic growth factors, thereby producing activated
CC   CTLs which are contacted with a carrier to form a composition. The
CC   composition can then be administered to the patient. The activated CTLs
CC   can be used for treating cancers, immune disorders, viral infections,
CC   AIDS, hepatitis, bacterial infection, fungal infection, malaria or
CC   tuberculosis.
XX
SQ   Sequence   9 AA;


  Query Match             100.0%;  Score 45;  DB 19;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 YLSGANLNL 9
             |||||||||
Db         1 ylsganlnl 9


RESULT   4
Y47655
ID   Y47655 standard; Peptide; 9 AA.
XX
AC   Y47655;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2266.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
```

```
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 118; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;


  Query Match             100.0%;  Score 45;  DB 20;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 YLSGANLNL 9
             |||||||||
Db         1 ylsganlnl 9


RESULT   5
Y09525
ID   Y09525 standard; peptide; 9 AA.
XX
AC   Y09525;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist CAP-1.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
```

```
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 1; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             100.0%;  Score 45;  DB 20;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANLNL 9
           |||||||||
Db       1 ylsganlnl 9

RESULT   6
Y09528
ID   Y09528 standard; peptide; 9 AA.
XX
AC   Y09528;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:4.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
```

```
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             95.6%;  Score 43;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANLNL 9
           ||||||:||
Db       1 ylsganinl 9

RESULT   7
Y54173
ID   Y54173 standard; peptide; 9 AA.
XX
AC   Y54173;
XX
DT   06-APR-2000  (first entry)
XX
DE   HLA binding peptide 1233.11 derived from source CEA.605V9.
XX
KW   Allele-specific binding motif; major histocompatibility complex; MHC;
KW   HLA; HLA-A2.1; cytotoxic T cell response; antigen; microbial infection;
KW   hepatitis; AIDS; malaria; condyloma acuminatum; cancer; prostate cancer;
KW   renal carcinoma; cervical carcinoma; lymphoma; tumour.
XX
OS   Unidentified.
XX
PN   WO9965522-A1.
XX
PD   23-DEC-1999.
XX
PF   17-JUN-1999;   99WO-US13789.
XX
PR   17-JUN-1998;   98US-0098584.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Sidney J,  Southwood S;
XX
DR   WPI; 2000-106018/09.
XX
PT   Novel HLA binding immunogenic peptides used to induce T cell activation
PT   and to induce an immune response  -
XX
PS   Claim 1; Page 32; 42pp; English.
XX
CC   Peptides Y54171-Y54236 represent immunogenic peptides comprising an
CC   allele-specific binding motif for the major histocompatibility complex
CC   (MHC) molecule HLA, e.g. HLA-A2.1. The peptides have conserved residues
CC   at certain positions such as positions 2 and 9. Also, the peptides do not
CC   comprise negative binding residues at other positions, such as positions
CC   1, 3, 6 and/or 7 (peptides 9 amino acids long) and at poisons 1, 3, 4,
CC   5, 7, 8 and/or 9 (peptides 10 amino acids long). The peptides are used to
CC   induce a cytotoxic T cell response to a preselected antigen. The method
CC   comprises contacting cytotoxic T cells from a patient (optionally
CC   expressing a specific MHC class I allele) with the present peptides.
CC   The peptides are used to treat and prevent microbial infection (e.g. in
CC   viral hepatitis B and C, human papilomavirus (HPV) infection, AIDS,
CC   cytomegalovirus (CMV), malaria, and condyloma acuminatum) and cancer
CC   (e.g. prostate cancer, renal carcinoma, cervical carcinoma, lymphoma).
CC   Patients in the acute phase of infection can be treated with the
CC   peptides in conjunction with other treatments. The antigenic peptides
CC   may be used to elicit cytotoxic T lymphocytes (CTLs) ex vivo and in
CC   vivo. The resulting CTLs can be used to treat chronic infections (viral
```

```
CC   or bacterial) or tumours in patients that do not respond to conventional
CC   forms of therapy. The peptides may also be used to produce monoclonal
CC   antibodies, which are useful as potential diagnostic or therapeutic
CC   agents. The peptides may also be used as diagnostic reagents.
XX
SQ   Sequence   9 AA;


  Query Match             93.3%;  Score 42;  DB 21;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGANLNL 9
             ||||||||:
Db       1 ylsganlnv 9


RESULT   8
Y09526
ID   Y09526 standard; peptide; 9 AA.
XX
AC   Y09526;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:2.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;


  Query Match             88.9%;  Score 40;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 YLSGANLNL 9
             |||||:|||
Db       1 ylsgadlnl 9


RESULT   9
W00680
ID   W00680 standard; peptide; 9 AA.
XX
AC   W00680;
XX
DT   01-MAY-1997  (first entry)
XX
DE   Peptide comprising residues 571-579 of Carcinoembryonic antigen.
XX
KW   Carcinoembryonic; antigen; human; cytotoxic T cell; pox virus;
KW   vector; epitope; determination; screening; tumour; treatment.
XX
OS   Homo sapiens.
XX
PN   WO9626271-A1.
XX
PD   29-AUG-1996.
XX
PF   13-FEB-1996;   96WO-US02156.
XX
PR   22-FEB-1995;   95US-0396385.
XX
PA   (THER-) THERION BIOLOGICS CORP.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Panicali D,  Schlom J,  Tsang KY;
XX
DR   WPI; 1996-402364/40.
XX
PT   Generation of human cytotoxic T-cells specific for CEA - useful in
PT   therapy, epitope mapping and drug screening
XX
PS   Claim 4; Page 57; 76pp; English.
XX
CC   Producing carcinoembryonic antigen (CEA) specific human cytotoxic T
CC   cells (CTC), comprises introducing a 1st pox virus vector, having
CC   at least 1 insertion site containing a DNA segment encoding a CEA
CC   peptide (i.e. the present peptide) to a host to stimulate CTC
CC   production, and at least 1 periodic interval after that, contacting
CC   the host with an additional antigen. The CEA specific CTC can be
CC   used to determine the CTC eliciting epitope of CEA, and to screen
CC   for compounds which enhance the ability of the antigen to create a
CC   CTC response. A host with a CEA expressing tumour can be treated by
CC   introducing the CTC to the host, and at least 1 periodic interval
CC   after that introducing a CEA peptide, i.e. the present peptide.
CC   The present peptide is positive for binding to HLA-A2, and scored
CC   561 and 806 in T2 cell binding assays, where the binding of an
CC   appropriate peptide results in the upregulation of surface HLA-A2
CC   on the T2 cells, which can be quantified via FACScan using an
CC   anti-HLA-A2 antibody (background 280 and 300).
XX
SQ   Sequence   9 AA;


  Query Match             84.4%;  Score 38;  DB 17;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 LSGANLNL 9
             ||||||||
Db       2 lsganlnl 9


RESULT  10
Y09527
ID   Y09527 standard; peptide; 9 AA.
```

```
XX
AC   Y09527;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:3.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             84.4%;  Score 38;  DB 20;  Length 9;
  Best Local Similarity   77.8%;  Pred. No. 2.1e+05;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSGANLNL 9
            |||||::||
Db        1 ylsgadinl 9

RESULT  11
Y09529
ID   Y09529 standard; peptide; 9 AA.
XX
AC   Y09529;
XX
DT   20-JUL-1999  (first entry)
XX
DE   Carcinoembryonic antigen peptide agonist SEQ ID NO:5.
XX
KW   Carcinoembryonic antigen; CEA; human; agonist; antagonist;
KW   immune response; carcinoma; gastrointestinal; breast; pancreatic;
KW   bladder; ovarian; lung; prostatic; T cell proliferation; cancer;
KW   adoptive transfer therapy; autoimmune reaction; immunotherapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9919478-A1.
XX
PD   22-APR-1999.
XX
PF   22-SEP-1998;   98WO-US19794.
XX
PR   10-OCT-1997;   97US-0061589.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Barzaga E,  Schlom J,  Zaremba S;
XX
DR   WPI; 1999-326544/27.
XX
PT   Peptide agonists and antagonists of carcinoembryonal antigen
XX
PS   Claim 5; Page 53; 72pp; English.
XX
CC   The present invention describes peptides (A) that comprise agonists (Ia)
CC   or antagonists (Ib) of human carcinoembryonal antigen (CEA). (Ia) are
CC   used in vaccines to kill or inhibit carcinoma cells that express CEA or
CC   its epitopes, particularly for treating gastrointestinal, breast,
CC   pancreatic, bladder, ovarian, lung or prostatic carcinoma. They can also
CC   be used to proliferate T cells, e.g. from vaccinated subjects, for use
CC   in adoptive transfer therapy. (Ib) are used to inhibit CEA-specific
CC   immune responses, e.g. in vaccinated subjects, to prevent an autoimmune
CC   reaction to cancer immunotherapy (i.e. to prevent attack on normal but
CC   CEA-expressing cells). (Ia) are more active than native sequence (I) and
CC   generate a highly specific and systemic anti-CEA response. Cytotoxic T
CC   cells generated recognize both (Ia) and native CEA epitopes. The present
CC   sequence represents a specifically claimed example of (Ia).
XX
SQ   Sequence   9 AA;

  Query Match             80.0%;  Score 36;  DB 20;  Length 9;
  Best Local Similarity   88.9%;  Pred. No. 2.1e+05;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGANLNL 9
            ||||| |||
Db        1 ylsgaclnl 9

RESULT  12
W00690
ID   W00690 standard; peptide; 9 AA.
XX
AC   W00690;
XX
DT   01-MAY-1997  (first entry)
XX
DE   NCA analogue of residues 571-579 of carcinoembryonic antigen.
XX
KW   Carcinoembryonic; antigen; epitope; NCA; analogue.
XX
OS   Homo sapiens.
XX
PN   WO9626271-A1.
XX
PD   29-AUG-1996.
XX
PF   13-FEB-1996;   96WO-US02156.
XX
PR   22-FEB-1995;   95US-0396385.
XX
PA   (THER-) THERION BIOLOGICS CORP.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
```

```
PI   Panicali D,  Schlom J,  Tsang KY;
XX
DR   WPI; 1996-402364/40.
XX
PT   Generation of human cytotoxic T-cells specific for CEA - useful in
PT   therapy, epitope mapping and drug screening
XX
PS   Example 2; Page 60; 76pp; English.
XX
CC   The present peptide is negative for binding to HLA-A2, and scored
CC   252 and 225 in T2 cell binding assays, where the binding of an
CC   appropriate peptide results in the upregulation of surface HLA-A2
CC   on the T2 cells, which can be quantified via FACScan using an
CC   anti-HLA-A2 antibody (background 280 and 300).
XX
SQ   Sequence   9 AA;


  Query Match             64.4%;  Score 29;  DB 17;  Length 9;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    6;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        1 YLSGANLNL 9
            |  | ||||
Db        1 yrpgenlnl 9


RESULT  13
W40267
ID   W40267 standard; Protein; 9 AA.
XX
AC   W40267;
XX
DT   16-JUN-1998  (first entry)
XX
DE   K. oxytoca R-specific amidohydrolase peptide T5.
XX
KW   R-specific amidohydrolase; hydrolysis; racemate; nitrogen source;
KW   3,3,3-trifluoro-2-hydroxy-2-methyl propionamide; isoform.
XX
OS   Klebsiella oxytoca.
XX
PN   WO9801568-A2.
XX
PD   15-JAN-1998.
XX
PF   10-JUL-1997;   97WO-EP03670.
XX
PR   03-MAR-1997;   97CH-0000500.
PR   10-JUL-1996;   96CH-0001723.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Brieden W,  Naughton A,  Robins K,  Shaw N,  Tinschert A;
PI   Zimmermann T;
XX
DR   WPI; 1998-101063/09.
XX
PT   Tri:fluoro-2-hydroxy-2-methyl-propionic acid enantiomer preparation
PT   - by stereoselective hydrolysis of corresponding racemic amide using
PT   microorganism or derived enzyme, used as drug intermediate
XX
PS   Example 10.2; Page 29; 68pp; German.
XX
CC   W40264-W40275 are peptide fragments of an R-specific amidohydrolase
CC   isolated from Klebsiella oxytoca strain PRS1. This enzyme allows the
CC   microorganism to utilise 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
CC   as its sole nitrogen source. This amidohydrolase is used in a process for
CC   preparing (R)-isoforms of 3,3,3-trifluoro-2-hydroxy-2-methyl propionamide
CC   which is cheaper than prior art optical resolution of the racemate using
CC   dimethoxy strychnine or (S)-(-)- alpha -methylbenzylamine.
XX
SQ   Sequence   9 AA;
```

```
  Query Match             48.9%;  Score 22;  DB 19;  Length 9;
  Best Local Similarity   62.5%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        1 YLSGANLN 8
            |  || ||
Db        1 ytvgamln 8


RESULT  14
R09414
ID   R09414 standard; Peptide; 6 AA.
XX
AC   R09414;
XX
DT   23-AUG-1990  (first entry)
XX
DE   LFA-1 alpha subunit polypeptide (k).
XX
KW   Lymphocyte function associated antigen; inflammation; metastasis.
XX
PN   EP362526-A.
XX
PD   11-APR-1990.
XX
PF   17-AUG-1989;   89EP-0115160.
XX
PR   23-AUG-1988;   88US-0235227.
PR   09-MAR-1989;   89US-0321017.
XX
PA   (DANA-) DANA FARBER CANCER.
XX
PI   Springer TA,  Larson R;
XX
DR   WPI; 1990-108985/15.
XX
PT   Pure alpha subunit of lymphocyte function associated antigen -
PT   and encoding DNA sequences, useful eg for suppressing
PT   inflammation or metastasis
XX
PS   Claim 6; Page 19; 27pp; English.
XX
CC   The alpha-subunit (a-SU), contg. al least one of the polypeptides
CC   given in R09404-417, can bind to ICAM-1 (or other natural
CC   ligands) on the surface of cells, and can associate with the beta-SU
CC   to form a heterodimer (also able to bind to ICAM-I).  a-SU, and its
CC   derivs., are useful in suppressing inflammation, metastasis and
CC   growth of a-SU expressing tumour cells and is used in the treatment
CC   of viral infections.
CC   The pref. dose is 1 pg - 10 mg/kg.
XX
SQ   Sequence   6 AA;


  Query Match             46.7%;  Score 21;  DB 11;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 YLSG 4
            ||||
Db        2 ylsg 5


RESULT  15
R88476
ID   R88476 standard; peptide; 6 AA.
XX
AC   R88476;
XX
DT   30-AUG-1996  (first entry)
XX
```

```
DE   Internal tryptic peptide from Tre6P synthase (peak 29) #2.
XX
KW   Tryptic peptide; trehalose-6-phosphate synthase; M. smegmatis; probe;
KW   trehalose; transgenic plant; heparin-activated; preservation; food;
KW   antigenic determinant; yeast; TSP1; fruit; berry; puree; jelly; jam.
XX
OS   Mycobacterium smegmatis.
XX
PN   WO9600789-A1.
XX
PD   11-JAN-1996.
XX
PF   29-JUN-1995;   95WO-FI00377.
XX
PR   29-JUN-1994;   94FI-0003133.
XX
PA   (ALKO-) ALKO GROUP LTD.
XX
PI   Holmstrom K,  Londesborough J,  Mandal A,  Mantyla E;
PI   Palva ET,  Tunnela O, Welin B;
XX
DR   WPI; 1996-077499/08.
XX
PT   New transgenic plants with increase trehalose contents - prepd. by
PT   transforming plants with a trehalose-6-phosphate synthase gene fused
PT   to a non-constitutive promoter
XX
PS   Example 6; Page 36; 55pp; English.
XX
CC   The sequences given in R88473-80 are internal tryptic peptides
CC   derived from trehalose-6-phosphate (Tre6P) synthase from M.
CC   smegmatis.  Tre6P is the key enzyme in the synthesis of trehalose
CC   via Tre6P.  The aim of the invention is to produce a transgenic
CC   plant with increase trehalose content.  Tre6P in M. smegmatis is
CC   heparin-activated and was isolated  and purified.    These peptides
CC   were derived from a protein which was purified with a mol. wt. of 55
CC   kD which  shared antigenic determinants with the yeast Tre6P
CC   synthase protein.  Using these peptides probes may be designed for
CC   the isolation of the Tre6P gene (TSP1) for the production of the
CC   transgenic plants.  The trehalose may be isolated from the transgenic
CC   plants and used in bulk preparation including the preservation of the
CC   flavour and structure of food stuffs during drying.  Fruits and berries
CC   form the transformed plants may be processed into purees, jellies and
CC   jams which have a fresher and richer flavour due to the increased
CC   trehalose content.
XX
SQ   Sequence   6 AA;


  Query Match             46.7%;  Score 21;  DB 17;  Length 6;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 YLSGA 5
            || ||
Db        1 ylega 5

RESULT  16
Y23411
ID   Y23411 standard; Peptide; 6 AA.
XX
AC   Y23411;
XX
DT   02-SEP-1999  (first entry)
XX
DE   V beta 6 clone found in MS patients after vaccination with TCR.
XX
KW   Vaccine; T cell receptor; TCR; T cell; V beta 6.2/3; V beta 6/5;
KW   V beta 6.7; V beta 2; V beta 5/1; V beta 7; V beta 13; V beta 6;
KW   multiple sclerosis.
XX
OS   Synthetic.
```

```
OS   Homo sapiens.
XX
PN   WO9927957-A1.
XX
PD   10-JUN-1999.
XX
PF   03-DEC-1997;   97WO-US23147.
XX
PR   03-DEC-1997;   97WO-US23147.
XX
PA   (IMMU-) IMMUNE RESPONSE CORP.
PA   (KIMM-) KIMMEL CANCER CENT SIDNEY.
XX
PI   Brostoff SW,  Carlo DJ,  Gold DP,  Smith LR,  Wilson DB;
XX
DR   WPI; 1999-404801/34.
XX
PT   T0 cell receptor peptide-derived vaccines
XX
PS   Example 11; Page 85; 104pp; English.
XX
CC   The specification describes vaccines which comprise immunologically
CC   effective amounts of T cell receptor (TCR) peptides. The TCRs are
CC   present on the surface of T cells. The TCRs are chosen from V beta 6.2/3,
CC   V beta 6/5, V beta 6.7, V beta 2, V beta 5/1, V beta 7 or V beta 13.
CC   The V beta TCR peptide-based vaccines are useful for prevention or
CC   treatment of multiple sclerosis (MS). The presence of V beta 6.7 appears
CC   to be particularly associated with multiple sclerosis and can be used
CC   to determine an individual's susceptibility to multiple sclerosis.
CC   Vaccinating, rather than passively administering heterologous antibodies,
CC   allows the host's own immune system to mobilize and suppress auto
CC   aggressive T cells.  Therefore, the suppression is persistent and may
CC   involve any and all immunological mechanisms in effecting that
CC   suppression.  Such a multi-faceted response is more effective than
CC   the uni-dimensional suppression achieved by passive administration of
CC   monoclonal antibodies or extant-derived regulatory T cell clones.
CC   Y23387-Y23480 represent peptides derived from TCR V beta 6 clones
CC   found in the cerebrospinal fluid (CFS) of MS patients, after vaccination
CC   with V beta 6.
XX
SQ   Sequence   6 AA;


  Query Match             46.7%;  Score 21;  DB 20;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        2 LSGANL 7
            |:| ||
Db        1 lagtnl 6


RESULT  17
R29031
ID   R29031 standard; Protein; 8 AA.
XX
AC   R29031;
XX
DT   20-APR-1993  (first entry)
XX
DE   Peptide encoded by ant-active toxin gene probe.
XX
KW   Toxin protein; ant.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1
FT                   /label= Arg or Lys
FT   Misc-difference 3
FT                   /label= Trp or Tyr
FT   Misc-difference 4
FT                   /label= Ile or Leu
```

```
FT   Misc-difference 7
FT                   /label= Ala or Asn
FT   Misc-difference 8
FT                   /label= Asn or Gln
XX
PN   WO9220802-A.
XX
PD   26-NOV-1992.
XX
PF   22-MAY-1992;   92WO-US04316.
XX
PR   22-MAY-1991;   91US-0703977.
PR   25-NOV-1991;   91US-0797645.
PR   12-MAY-1992;   92EP-0304228.
XX
PA   (MYCO ) MYCOGEN CORP.
XX
PI   Kennedy MK,  Meier H,  Payne JM,  Randall JB,  Uick HJ;
XX
DR   WPI; 1992-415780/50.
DR   N-PSDB; Q31414, Q31415.
XX
PT   Toxin proteins isolated from Bacillus thuringiensis - for controlling
PT   amts. e.g. fire, carpenter, argentine and pharaoh ants
XX
PS   Disclosure; Page 56; 71pp; English.
XX
CC   The peptide (or point mutation variants as shown in the features)
CC   is encoded by nucleotide probes used in the rapid identification of
CC   Bacillus thuringiensis ant-active toxin genes.
XX
SQ   Sequence   8 AA;


  Query Match             46.7%;  Score 21;  DB 13;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGAN 6
           :::|||
Db       3 wingan 8


RESULT  18
R78208
ID   R78208 standard; Peptide; 8 AA.
XX
AC   R78208;
XX
DT   22-FEB-1996  (first entry)
XX
DE   B. thuringiensis endotoxin derived peptide.
XX
KW   Endotoxin; probes; microbes; peptide; detection; nematode.
XX
OS   Bacillus thuringiensis.
XX
PN   US5430137-A.
XX
PD   04-JUL-1995.
XX
PF   25-OCT-1989;   89US-0427068.
XX
PR   30-OCT-1992;   92US-0968781.
PR   25-OCT-1989;   89US-0427068.
PR   26-JUL-1991;   91US-0737569.
XX
PA   (MYCO ) MYCOGEN CORP.
XX
PI   Gaertner FH,  Narva KE,  Schnepf HE,  Schwab GE,  Sick AJ;
PI   Thompson M;
XX
DR   WPI; 1995-245777/32.
```

```
XX
PT   Nucleotide sequence used as probes to identify Bacillus
PT   thuringiensis - are derived from the B.thuringiensis endotoxin
PT   genes, for identifying microbes which encode toxins
XX
PS   Claim 1; Columns 41-42; 30pp; English.
XX
CC   The nucleotide sequence which encodes R78208, a B. thuringiensis
CC   (B.t.) endotoxin derived peptide, specifies the degenerate
CC   probe Q94857. The probe can be used for the detection
CC   of endotoxin producing B.t. microbes. The probe aids in the search
CC   for useful microbes hosting toxin encoding genes, specifically from
CC   nematode species.
XX
SQ   Sequence   8 AA;

  Query Match             46.7%;  Score 21;  DB 16;  Length 8;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 YLSGAN 6
           :::|||
Db       3 wingan 8

RESULT  19
R37221
ID   R37221 standard; peptide; 9 AA.
XX
AC   R37221;
XX
DT   06-SEP-1993  (first entry)
XX
DE   IL-6 antagonist peptide #5.
XX
KW   IL-6; antagonist; cytokine; B cell differentiation; inflammation;
KW   tissue injury; B9.9 hybridoma cell line; Castleman's Disease;
KW   Lennart's T-cell lymphoma; Non-Hodgkin's lymphoma; allergy;
KW   immune deficiency disease; cardiac myxoma;
KW   mesangial proliferative glomerulonephritis.
XX
PN   US5210075-A.
XX
PD   11-MAY-1993.
XX
PF   16-FEB-1990;   90US-0480868.
XX
PR   16-FEB-1990;   90US-0480868.
XX
PA   (TANA ) TANABE SEIYAKU CO.
XX
PI   Chiang S,  Lobl TJ,  Nagarajan G,  Scholz W;
XX
DR   WPI; 1993-166970/20.
XX
PT   New peptide interleukin-6 antagonists - for treating and
PT   preventing auto-immune, immuno:inflammatory, neoplastic and
PT   infectious diseases etc.
XX
PS   Example 3; Column 27; 20pp; English.
XX
CC   This peptide is a specific example of a highly generic claimed
CC   formula covering sequences derived from the p51-70 portion (i.e the
CC   IL-6 receptor-binding portion) of IL-6 (see R37216) or modelled
CC   after different portions of this sequence. In an assay to determine
CC   IL-6 antagonist activity of this peptide, proliferation of the IL-6
CC   dependent B9.9 hybridoma cell line (J.Immunol. 139: 4116, 1987) was
CC   inhibited.  See R37216-R37261.
XX
SQ   Sequence   9 AA;
```

```
  Query Match             46.7%;  Score 21;  DB 14;  Length 9;
  Best Local Similarity   62.5%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LSGANLNL 9
           |:  ||||
Db       1 laennlnl 8


RESULT  20
R73808
ID   R73808 standard; peptide; 9 AA.
XX
AC   R73808;
XX
DT   19-JUN-1995  (first entry)
XX
DE   Antigen fragment 124, from LCMV has binding affinity for HLA-2.1.
XX
KW   antigen; epitope; immunogenic target protein; PSA; HBVc; HBVs; EBV;
KW   HIV1; plasma specific antigen; hepatitis B virus; Epstein Barr;
KW   human immunodeficiency virus; human papilloma virus; p53; c-ERB2;
KW   MAGE-1; melanoma antigen-1; core antigen; surface antigen;
KW   pharmaceutical composition; in vivo; ex vivo; therapeutic;
KW   diagnostic; MHC class I molecule; major histocompatability complex;
KW   HLA-A2.1; 9mer; 10mer; anchor; human leukocyte antigen; PLP; 8mer;
KW   algorithm prediction; MBP; CMV; cytomegalovirus; HSV;
KW   herpes simplex virus; influenza A; M1; LCMV.
XX
OS   Synthetic.
XX
PN   WO9420127-A.
XX
PD   15-SEP-1994.
XX
PF   04-MAR-1994;   94WO-US02353.
XX
PR   05-MAR-1993;   93US-0027146.
PR   04-JUN-1993;   93US-0073205.
PR   29-NOV-1993;   93US-0159184.
XX
PA   (CYTE-) CYTEL CORP.
XX
PI   Grey HM,  Kast WM,  Sette A,  Sidney J;
XX
DR   WPI; 1994-302678/37.
XX
PT   Immunogenic peptide(s) having an HLA-A2.1 binding motif - used
PT   for treatment or prophylaxis of cancer, virus infection or
PT   autoimmune diseases.
XX
PS   Disclosure; Page 84; 138pp; English.
XX
CC   R73685-876 are potential peptide binders of HLA-A2.1 motif. Using
CC   motifs disclosed in the invention, these peptides were screened for
CC   further motifs. Only peptides with binding affinity of at least 1%
CC   (binding affinity is expressed as an IC50 value) as compared to the
CC   standard peptide (R71293) in assays. This peptide from LCMV (sic) Np
CC   antigen has a binding value of 0.0280. The peptides of the invention
CC   can induce cytotoxic T lymphocytes which can react with target cells.
CC   They can be used for the treatment or prophlyaxis of cancer, eg.
CC   prostate cancer or lymphoma, etc.
XX
SQ   Sequence   9 AA;

  Query Match             46.7%;  Score 21;  DB 15;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LSGANLNL 9
           :|| |  :|
Db       2 isgynfsl 9
```

```
RESULT  21
W54298
ID   W54298 standard; Peptide; 9 AA.
XX
AC   W54298;
XX
DT   28-AUG-1998  (first entry)
XX
DE   Human cytohesin-1 PH domain randomised AB loop (clone 7).
XX
KW   Pleckstrin homology domain; PH domain; peptide library;
KW   functional molecular surface; protein structural template;
KW   vaccine; gene therapy; cytohesin 1; human.
XX
OS   Synthetic.
XX
PN   WO9745538-A1.
XX
PD   04-DEC-1997.
XX
PF   30-MAY-1997;   97WO-EP02840.
XX
PR   31-MAY-1996;   96EP-0108776.
XX
PA   (MEDI-) MEDIGENE AG.
XX
PI   Bruhn H,  Funk M,  Henkel T,  Steipe B;
XX
DR   WPI; 1998-230215/20.
DR   N-PSDB; V26501.
XX
PT   Vectors used to produce PH domain-like peptide libraries - which are
PT   screened for therapeutically useful peptide(s), e.g. to produce
PT   vaccines
XX
PS   Example 5; Fig 12; 137pp; English.
XX
CC   This is a randomised AB loop encoded by clone 7 (see V26501) of a
CC   large peptide library (2 million clones) of synthetic Pleckstrin
CC   homology (PH) domains with randomised discontinuous surface
CC   epitopes.  The progenitor AB sequence comprises amino acid residues
CC   16-21 of a stabilised synthetic PH domain (see W54310) of human
CC   cytohesin 1.  The peptide library was generated by randomisation of
CC   AB and CD loop regions using randomised oligonucleotide primers
CC   (see W26493-94).  The randomised AB and CD loop regions of 9 clones
CC   (see W54292-309) of the peptide library are provided.  Randomisation
CC   did not compromise the structural integrity and folding stability of
CC   the progenitor domain.  The invention provides vectors that are used
CC   for the production of PH domain-like peptide libraries, which can
CC   be screened to identify peptides that have desirable properties,
CC   especially novel binding or catalytic properties, and which may
CC   be of use in research or therapy, or as vaccines.  Novel synthetic
CC   protein structural templates for the generation, screening and
CC   evolution of functional molecular surfaces are provided.
XX
SQ   Sequence   9 AA;

  Query Match             46.7%;  Score 21;  DB 19;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 LSGANLN 8
           |||  :|
Db       1 lsggrvn 7


RESULT  22
Y41847
ID   Y41847 standard; Peptide; 7 AA.
XX
```

```
AC   Y41847;
XX
DT   09-DEC-1999  (first entry)
XX
DE   Rheumatoid arthritis diagnostic protein isoform 17 peptide #2.
XX
KW   Human; rheumatoid arthritis; RA; diagnosis; RPI; RADF; detection;
KW   rheumatoid arthritis diagnostic feature; ERPI; synovial fluid;
KW   rheumatoid arthritis diagnostic protein isoform; screening;
KW   expression reference protein isoform; prognosis.
XX
OS   Homo sapiens.
XX
PN   WO9947925-A2.
XX
PD   23-SEP-1999.
XX
PF   15-MAR-1999;   99WO-GB00763.
XX
PR   13-MAR-1998;   98GB-0005477.
XX
PA   (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI   Parekh RB,  Patel TP,  Townsend RR;
XX
DR   WPI; 1999-571871/48.
XX
PT   Diagnosis of human rheumatoid arthritis by two-dimensional
PT   electrophoresis -
XX
PS   Claim 20; Page 150; 157pp; English.
XX
CC   A method has been developed for the diagnosis of human rheumatoid
CC   arthritis (RA) using two-dimensional electrophoresis to generate a
CC   two-dimensional array of features. The method can be used for screening,
CC   diagnosis and prognosis of RA in a subject or for monitoring the effect
CC   of an anti-RA drug or therapy administered to a subject. The method
CC   comprises: (a) analysing a sample of serum or plasma and optionally
CC   synovial fluid by two-dimensional electrophoresis, to generate a two-
CC   dimensional array of features; (b) identifying at least one chosen
CC   feature whose relative abundance correlates with the presence or absence
CC   of RA; and (c) comparing the abundance of each chosen feature in the
CC   sample with the abundance of that chosen feature in serum or plasma from
CC   one or more persons without RA, where the relative abundance of the
CC   chosen feature or features in the sample indicates the presence or
CC   absence of RA in the subject. The method can also be used in clinical
CC   studies for testing drugs for therapy of RA, for purification of RA-
CC   diagnostic protein isoforms (RPIs), and for production of antibodies to
CC   RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify
CC   compounds that promote or inhibit their activity, which are then used as
CC   RA drugs. Nucleic acid encoding RADFs can be used in gene therapy
CC   protocols. Y41844 to Y42100 represent RPI peptides, Y42101 to Y42103
CC   represent expression reference protein isoform peptides and Z25066 to
CC   Z25068 represent degenerate probes for RPIs, which are all used in
CC   the exemplification of the present invention.
XX
SQ   Sequence   7 AA;


  Query Match             44.4%;  Score 20;  DB 20;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        3 SGANLN 8
            |||:|:
Db        2 sgadls 7


RESULT  23
R34629
ID   R34629 standard; Protein; 8 AA.
XX
AC   R34629;
```

```
XX
DT   05-AUG-1993  (first entry)
XX
DE   B. thuringiensis dipteran toxin probe (m) design peptide.
XX
KW   Probe; Bacillus thuringiensis; Bt; insecticidal; active; endotoxin;
KW   lepidoptera; diptera; coleoptera.
XX
OS   Synthetic.
XX
PN   US5204237-A.
XX
PD   20-APR-1993.
XX
PF   25-OCT-1989;   89US-0427068.
XX
PR   25-OCT-1989;   89US-0427068.
PR   26-JUL-1991;   91US-0737569.
XX
PA   (MYCO ) MYCOGEN CORP.
XX
PI   Gaertner FH,  Schwab GE,  Sick AJ;
XX
DR   WPI; 1993-143952/17.
XX
PT   Gene probe for identifying endotoxin sequences in Bacillus
PT   thuringiensis - are universal, type specific or gene specific,
PT   for rapid detection and characterisation of insecticidal activity
XX
PS   Disclosure; Column 2; 5pp; English.
XX
CC   The sequences given in R35090-100 and R34628-32 are peptides which
CC   were used to design the probes given in Q40219-34.  These probes can
CC   be used to identify Bacillus thuringiensis (Bt) DNA which encodes
CC   insecticidally active endotoxin.  Probe (a) identifies genes encoding
CC   a toxin effective against any sort of insect.  Sequences (b) and (e)-
CC   (l) are specific for toxins against lepidoptera; sequences (c), (m)
CC   and (n) for toxins active against diptera and sequence (d), (o) and
CC   (p) for toxins active against coleoptera.
XX
SQ   Sequence   8 AA;


  Query Match             44.4%;  Score 20;  DB 14;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        4 GANLN 8
            | |:|
Db        2 gtnmn 6


RESULT  24
R78218
ID   R78218 standard; Peptide; 8 AA.
XX
AC   R78218;
XX
DT   22-FEB-1996  (first entry)
XX
DE   B. thuringiensis endotoxin derived peptide.
XX
KW   Endotoxin; probe; microbes; peptide; detection; dipteran.
XX
OS   Bacillus thuringiensis.
XX
PN   US5430137-A.
XX
PD   04-JUL-1995.
XX
PF   25-OCT-1989;   89US-0427068.
XX
PR   30-OCT-1992;   92US-0968781.
```

```
PR   25-OCT-1989;   89US-0427068.
PR   26-JUL-1991;   91US-0737569.
XX
PA   (MYCO ) MYCOGEN CORP.
XX
PI   Gaertner FH,  Narva KE,  Schnepf HE,  Schwab GE,  Sick AJ;
PI   Thompson M;
XX
DR   WPI; 1995-245777/32.
XX
PT   Nucleotide sequence used as probes to identify Bacillus
PT   thuringiensis - are derived from the B.thuringiensis endotoxin
PT   genes, for identifying microbes which encode toxins
XX
PS   Claim 1; Columns 49-50; 30pp; English.
XX
CC   The nucleotide sequences which encode R78218 and R78219, B.
CC   thuringiensis (B.t.) endotoxin derived peptides, specify the probes
CC   Q94867 and Q94868 respectively. The probes can be used for the
CC   detection of endotoxin producing B.t. microbes. The probes aid in
CC   the search for useful microbes hosting toxin encoding genes,
CC   specifically from dipteran species.
XX
SQ   Sequence   8 AA;

  Query Match             44.4%;  Score 20;  DB 16;  Length 8;
  Best Local Similarity   60.0%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 GANLN 8
          | |:|
Db      2 gtnmn 6


RESULT  25
W36092
ID   W36092 standard; peptide; 8 AA.
XX
AC   W36092;
XX
DT   16-MAR-1998  (first entry)
XX
DE   E. coli DNA polymerase III chi subunit tryptic peptide chi-4.
XX
KW   Psi subunit; E. coli; DNA polymerase III holoenzyme; probe; enzyme;
KW   hybridisation; PCR.
XX
OS   Escherichia coli.
XX
PN   US5668004-A.
XX
PD   16-SEP-1997.
XX
PF   22-JUL-1994;   94US-0279058.
XX
PR   22-JUL-1994;   94US-0279058.
PR   22-JAN-1993;   93US-0826926.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   O'donnell ME;
XX
DR   WPI; 1997-470038/43.
XX
PT   Isolated theta, chi, psi, delta and delta' protein sub:units of E.
PT   coli polymerase III holo:enzyme - used to make man-made enzyme of 5
PT   or 6 sub:units, useful for long chain PCR
XX
PS   Example 3; Column 57-58; 65pp; English.
XX
CC   This is the amino acid sequence of a tryptic peptide from the chi subunit
CC   of the E. coli DNA polymerase III holoenzyme (W36068).  Probes derived
```

```
CC   from the tryptic peptides W36089-W36092 were used to isolate the chi
CC   subunit gene sequence from a lambda phage library.
CC   The E. coli polymerase III subunits (theta, chi, psi, delta and delta'
CC   (T98151-T98155 respectively)) are used to make man-made enzymes
CC   comprising 5 or 6 subunits and potentially for use in long chain PCR.
XX
SQ   Sequence   8 AA;

  Query Match             44.4%;  Score 20;  DB 18;  Length 8;
  Best Local Similarity   80.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 GANLN 8
          | |||
Db      1 gfnln 5


RESULT  26
R69993
ID   R69993 standard; peptide; 9 AA.
XX
AC   R69993;
XX
DT   19-OCT-1995  (first entry)
XX
DE   Nonameric mimotope 24 used to obtain highly specific antibodies.
XX
KW   mimotope; antibody; production; high specificity; detection;
KW   immunoassay; high performance liquid chromatography.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   9
FT                   /note= "Arg-NH2"
XX
PN   US5384263-A.
XX
PD   24-JAN-1995.
XX
PF   13-OCT-1987;   87US-0108130.
XX
PR   11-OCT-1988;   88US-0255906.
PR   13-OCT-1987;   87US-0108130.
PR   04-JUN-1993;   93US-0072190.
XX
PA   (TERR-) TERRAPIN TECHNOLOGIES INC.
XX
PI   Kauvar LM;
XX
DR   WPI; 1995-105497/14.
XX
PT   Producing antibodies with high specificity and affinity for an
PT   analyte  - by immunisation with selected mimotope, also analyte
PT   detection kits, useful for immunoassay of materials usually
PT   analysed by HPLC
XX
PS   Example 5; Fig 6; 25pp; English.
XX
CC   R69970-993 are nonameric mimotopes designed to show high diversity
CC   in hydrophobic moment and hydrophobic index, as well as charge
CC   distribution and size. The mimotopes are used in the method of the
CC   invention to obtain antibodies specifically and strongly reactive
CC   with a desired analyte. 16 of the peptides were tested for ability
CC   to bind the murine antibody Mab33-6, arbitrarily chosen, and known
CC   to bind to the peptides MB3 and MB4. 3 of the 16 peptides successfully
CC   bound Mab33-6. The mimotope is obtd. by reacting a panel of starting
CC   antibodies (Abs) representative of the resting B cell repertoire of a
CC   mammal with an analyte (so as to identify analyte-reacting Abs) and
CC   then reacting each of a panel of candidate mimotopes representative
CC   of a random set of 3D contours with the analyte-reacting Abs. A
CC   subject is immunised with one or more mimotopes identified and the
```

```
CC   product Abs are recovered from the serum of the subject.
XX
SQ   Sequence   9 AA;


  Query Match             44.4%;  Score 20;  DB 16;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 SGAN 6
           ||||
Db       5 sgan 8


RESULT  27
W38383
ID   W38383 standard; peptide; 9 AA.
XX
AC   W38383;
XX
DT   08-APR-1998  (first entry)
XX
DE   Synthetic pMEL17 peptide.
XX
KW   Melanoma; immunogen; cytotoxic T lymphocyte; CTL;
KW   human leukocyte antigen-A1; HLA-A1; human leukocyte antigen-A3;
KW   HLA-A3; epitope; pMEL-17; tyrosinase; vaccine; protection.
XX
OS   Synthetic.
XX
PN   WO9734613-A1.
XX
PD   25-SEP-1997.
XX
PF   17-MAR-1997;   97WO-US04958.
XX
PR   04-OCT-1996;   96US-0027627.
PR   19-MAR-1996;   96US-0013972.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI   Cox AL,  Engelhard VH,  Hendrikson RC,  Hunt DF,  Kittlesen D;
PI   Shabanowitz J,  Skipper J,  Slingluff CL;
XX
DR   WPI; 1997-479982/44.
XX
PT   Melanoma-specific immunogens of pMel-17 and tyrosinase - useful in
PT   vaccination for producing melanoma-specific cytotoxic T lymphocytes
XX
PS   Example 9; Page 65; 106pp; English.
XX
CC   The present peptide was used in the preparation of a novel melanoma
CC   specific immunogen, comprising at least 1 melanoma specific
CC   cytotoxic T lymphocyte (CTL) epitope, where at least 1 of the
CC   epitopes is substantially homologous to a human leukocyte
CC   antigen-A1 (HLA-A1) and HLA-A3 restricted epitope of a melanoma
CC   antigen, either pMEL-17 or tyrosinase. The immunogen can be used in
CC   vaccines for protection against melanoma in mammals.
XX
SQ   Sequence   9 AA;


  Query Match             44.4%;  Score 20;  DB 18;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANLN 8
             ||: |:|:
Db       1 ylaeadls 8


RESULT  28
W19861
ID   W19861 standard; peptide; 9 AA.
XX
AC   W19861;
XX
DT   24-FEB-1998  (first entry)
XX
DE   Fragment of enzyme capable of catalysing conversion of DNA to CO-DNA.
XX
KW   CO-DNA; cell division; cell proliferation; tumour; lysilendopeptidase;
KW   N-terminal fragment; V8 proteinase digestion.
XX
OS   Gallus domesticus.
XX
PN   WO9720919-A1.
XX
PD   12-JUN-1997.
XX
PF   27-NOV-1996;   96WO-IB01323.
XX
PR   06-DEC-1995;   95US-0567934.
XX
PA   (KAGA/) KAGAWA H.
PA   (KAGA/) KAGAWA K.
PA   (TOKI/) TOKIMATSU H.
XX
PI   Kagawa H,  Kagawa K,  Tokimatsu H;
XX
DR   WPI; 1997-319767/29.
XX
PT   New enzyme which converts DNA to CO-DNA - used to reduce or
PT   eliminate the ability of a cell to divide, for inhibiting cell
PT   proliferation and treating tumours
XX
PS   Claim 3; Page 17; 35pp; English.
XX
CC   This sequence represents a V8 proteinase digestion product of an
CC   isolated pure enzyme which is capable of catalysing the conversion of
CC   DNA to CO-DNA. CO-DNA is a form of DNA in which a carbonyl group is
CC   attached to the 1' carbon of the sugar constituent of the DNA. The enzyme
CC   can be used to convert DNA to CO-DNA which reduces or eliminates the
CC   ability of a cell to divide. The enzyme can be used for inhibiting cell
CC   proliferation and for treating tumours.
XX
SQ   Sequence   9 AA;


  Query Match             44.4%;  Score 20;  DB 18;  Length 9;
  Best Local Similarity   42.9%;  Pred. No. 2.1e+05;
  Matches    3;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 LSGANLN 8
             : |||::
Db       3 idganvh 9


RESULT  29
Y55448
ID   Y55448 standard; peptide; 9 AA.
XX
AC   Y55448;
XX
DT   17-JAN-2000  (first entry)
XX
DE   HLA binding plu-1 peptide.
XX
KW   Human; cancer-associated polypeptide; plu-1; ovarian cancer; vaccine;
KW   breast cancer; immune response; cytotoxic T lymphocyte; imaging agent;
KW   therapeutic; plu-1 antigen; breast tumor; tumor antigen; epitope.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9949034-A1.
```

```
XX
PD   30-SEP-1999.
XX
PF   19-MAR-1999;   99WO-GB00866.
XX
PR   20-MAR-1998;   98GB-0005877.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Taylor-papadimitriou J;
XX
DR   WPI; 1999-591090/50.
XX
PT   New nucleic acid encoding the cancer-associated polypeptide plu-1, for
PT   diagnosis, treatment and prevention of cancer, especially of breast and
PT   ovary  -
XX
PS   Example 2; Fig 12; 173pp; English.
XX
CC   The invention relates to a human cancer-associated polypeptide plu-1. The
CC   plu-1 polypeptide can be recombinantly expressed by standard recombinant
CC   methodology. Detection of the plu-1 nucleic acid or the polypeptide is
CC   used for the following: (i) diagnosis (including imaging) and prognosis
CC   of, and determination of susceptibility to, cancer, specifically ovarian
CC   or breast cancer; and ii) treating cancer (by inducing an immune response
CC   against cancer cells, e.g. as a vaccine, or by antisense inhibition).
CC   Antigens derived from the polypeptide are used to generate activated
CC   cytotoxic T lymphocytes, or dendritic cells, for subsequent return to
CC   the patient for treatment of cancer. The polypeptide may also be used to
CC   identify inhibitors of plu-1 activity. Fragments of the polypeptide, and
CC   antibodies raised against plu-1, are useful as assay and imaging agents,
CC   also therapeutically (to induce an anti-idiotype response or where
CC   conjugated to cytotoxic agents). The plu-1 antigen is expressed more
CC   commonly in breast tumors than some known tumor antigens. Sequences
CC   Y55320-629 represent predicted peptides from the plu-1 polypeptide which
CC   may bind to the human class I alleles B27, A2, A3 and A11.
XX
SQ   Sequence   9 AA;


  Query Match             44.4%;  Score 20;  DB 20;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANL 7
           :||| :|
Db       3 flsgdsl 9


RESULT  30
Y55529
ID   Y55529 standard; peptide; 9 AA.
XX
AC   Y55529;
XX
DT   17-JAN-2000  (first entry)
XX
DE   HLA binding plu-1 peptide.
XX
KW   Human; cancer-associated polypeptide; plu-1; ovarian cancer; vaccine;
KW   breast cancer; immune response; cytotoxic T lymphocyte; imaging agent;
KW   therapeutic; plu-1 antigen; breast tumor; tumor antigen; epitope.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9949034-A1.
XX
PD   30-SEP-1999.
XX
PF   19-MAR-1999;   99WO-GB00866.
XX
PR   20-MAR-1998;   98GB-0005877.
```

```
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Taylor-papadimitriou J;
XX
DR   WPI; 1999-591090/50.
XX
PT   New nucleic acid encoding the cancer-associated polypeptide plu-1, for
PT   diagnosis, treatment and prevention of cancer, especially of breast and
PT   ovary  -
XX
PS   Example 2; Fig 12; 173pp; English.
XX
CC   The invention relates to a human cancer-associated polypeptide plu-1. The
CC   plu-1 polypeptide can be recombinantly expressed by standard recombinant
CC   methodology. Detection of the plu-1 nucleic acid or the polypeptide is
CC   used for the following: (i) diagnosis (including imaging) and prognosis
CC   of, and determination of susceptibility to, cancer, specifically ovarian
CC   or breast cancer; and ii) treating cancer (by inducing an immune response
CC   against cancer cells, e.g. as a vaccine, or by antisense inhibition).
CC   Antigens derived from the polypeptide are used to generate activated
CC   cytotoxic T lymphocytes, or dendritic cells, for subsequent return to
CC   the patient for treatment of cancer. The polypeptide may also be used to
CC   identify inhibitors of plu-1 activity. Fragments of the polypeptide, and
CC   antibodies raised against plu-1, are useful as assay and imaging agents,
CC   also therapeutically (to induce an anti-idiotype response or where
CC   conjugated to cytotoxic agents). The plu-1 antigen is expressed more
CC   commonly in breast tumors than some known tumor antigens. Sequences
CC   Y55320-629 represent predicted peptides from the plu-1 polypeptide which
CC   may bind to the human class I alleles B27, A2, A3 and A11.
XX
SQ   Sequence   9 AA;


  Query Match             44.4%;  Score 20;  DB 20;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANL 7
           :||| :|
Db       2 flsgdsl 8


RESULT  31
Y55623
ID   Y55623 standard; peptide; 9 AA.
XX
AC   Y55623;
XX
DT   17-JAN-2000  (first entry)
XX
DE   HLA binding plu-1 peptide.
XX
KW   Human; cancer-associated polypeptide; plu-1; ovarian cancer; vaccine;
KW   breast cancer; immune response; cytotoxic T lymphocyte; imaging agent;
KW   therapeutic; plu-1 antigen; breast tumor; tumor antigen; epitope.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9949034-A1.
XX
PD   30-SEP-1999.
XX
PF   19-MAR-1999;   99WO-GB00866.
XX
PR   20-MAR-1998;   98GB-0005877.
XX
PA   (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX
PI   Taylor-papadimitriou J;
XX
DR   WPI; 1999-591090/50.
```

```
XX
PT   New nucleic acid encoding the cancer-associated polypeptide plu-1, for
PT   diagnosis, treatment and prevention of cancer, especially of breast and
PT   ovary  -
XX
PS   Example 2; Fig 12; 173pp; English.
XX
CC   The invention relates to a human cancer-associated polypeptide plu-1. The
CC   plu-1 polypeptide can be recombinantly expressed by standard recombinant
CC   methodology. Detection of the plu-1 nucleic acid or the polypeptide is
CC   used for the following: (i) diagnosis (including imaging) and prognosis
CC   of, and determination of susceptibility to, cancer, specifically ovarian
CC   or breast cancer; and ii) treating cancer (by inducing an immune response
CC   against cancer cells, e.g. as a vaccine, or by antisense inhibition).
CC   Antigens derived from the polypeptide are used to generate activated
CC   cytotoxic T lymphocytes, or dendritic cells, for subsequent return to
CC   the patient for treatment of cancer. The polypeptide may also be used to
CC   identify inhibitors of plu-1 activity. Fragments of the polypeptide, and
CC   antibodies raised against plu-1, are useful as assay and imaging agents,
CC   also therapeutically (to induce an anti-idiotype response or where
CC   conjugated to cytotoxic agents). The plu-1 antigen is expressed more
CC   commonly in breast tumors than some known tumor antigens. Sequences
CC   Y55320-629 represent predicted peptides from the plu-1 polypetide which
CC   may bind to the human class I alleles B27, A2, A3 and A11.
XX
SQ   Sequence   9 AA;


  Query Match             44.4%;  Score 20;  DB 20;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 YLSGANL 7
           :||| :|
Db       2 flsgdsl 8


RESULT  32
Y46533
ID   Y46533 standard; Peptide; 9 AA.
XX
AC   Y46533;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #1144.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
```

```
PS   Claim 1; Page 75; 150pp; English.
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;


  Query Match             44.4%;  Score 20;  DB 20;  Length 9;
  Best Local Similarity   57.1%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 YLSGANL 7
           |  | ||
Db       3 yrpgvnl 9


RESULT  33
Y47062
ID   Y47062 standard; Peptide; 9 AA.
XX
AC   Y47062;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #1673.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 92; 150pp; English.
```

```
XX
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;

  Query Match             44.4%;  Score 20;  DB 20;  Length 9;
  Best Local Similarity   50.0%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 YLSGANLN 8
              ||: |:|:
Db          1 ylaeadls 8


RESULT  34
Y47532
ID   Y47532 standard; Peptide; 9 AA.
XX
AC   Y47532;
XX
DT   01-DEC-1999  (first entry)
XX
DE   Immunogenic peptide having a human leukocyte antigen binding motif #2143.
XX
KW   Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW   immune response; T cell activation; major histocompatibility complex;
KW   cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW   prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW   vaccine; immunisation.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9945954-A1.
XX
PD   16-SEP-1999.
XX
PF   13-MAR-1998;   98WO-US05039.
XX
PR   13-MAR-1998;   98WO-US05039.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Sette A,  Kubo RT,  Sidney J,  Celis E,  Grey HM,  Southwood S;
XX
DR   WPI; 1999-551214/46.
XX
PT   New immunogenic peptides with HLA binding motif, useful in treatment
PT   and diagnosis of cancers and viral diseases  -
XX
PS   Claim 1; Page 113; 150pp; English.
XX
```

```
CC   Y45390 to Y48214 represent specifically claimed immunogenic peptides
CC   having a human major histocompatibility complex (MHC) Class I (also
CC   known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC   peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC   HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC   response against the antigen from which the peptide is derived.
CC   Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC   normally induced by an antigen in the form of a peptide fragment bound
CC   to a HLA molecule, rather than the intact foreign antigen itself, and
CC   are particularly important in tumour rejection and in fighting viral
CC   infections. The peptides are therefore useful therapeutically to treat
CC   or prevent viral infections and cancers in mammals (especially humans)
CC   e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC   They can be administered as vaccines to elicit an immune response in
CC   individuals susceptible or otherwise at risk of viral infection or
CC   cancer, or used to treat chronic or acute conditions. They are also
CC   useful diagnostically, and can be used to induce a cytotoxic T cell
CC   response, by contacting a cytotoxic T cell with the peptide e.g. to
CC   produce CTLs ex vivo for infusion back into a patient. The
CC   polynucleotides encoding the immunogenic peptides are also useful
CC   therapeutically and for immunisation as above.
XX
SQ   Sequence   9 AA;

  Query Match             44.4%;  Score 20;  DB 20;  Length 9;
  Best Local Similarity   55.6%;  Pred. No. 2.1e+05;
  Matches    5;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 YLSGANLNL 9
               :||  |||
Db          1 flskqylnl 9


RESULT  35
R46809
ID   R46809 standard; Protein; 7 AA.
XX
AC   R46809;
XX
DT   19-AUG-1994  (first entry)
XX
DE   Phytase derived peptide 675.
XX
KW   pH 2.5; acid phosphatase; Trichoderma; host; Aspergillus; phytic acid;
KW   phytate degrading enzyme; PDE; removal; inositol hexaphosphoric acid;
KW   plant; feed composition; filtration.
XX
OS   Synthetic.
XX
PN   WO9403612-A.
XX
PD   17-FEB-1994.
XX
PF   30-JUL-1993;   93WO-FI00310.
XX
PR   31-JUL-1992;   92US-0923724.
XX
PA   (ALKO-) ALKO LTD.
XX
PI   Cantrell M,  Fagerstroem RB,  Miettinen-Oinonen ASK;
PI   Nevalainen HK,  Paloheimo MT,  Piddington C,  Rambosek JA;
PI   Torkkeli TK,  Turunen MK;
XX
DR   WPI; 1994-065700/08.
XX
PT   Compsns. contg. phytate degrading enzymes - obtd. by expression
PT   of their genes in Trichoderma, used partic. for producing animal
PT   feed compsns.
XX
PS   Example 4; Page 43; 142pp; English.
XX
CC   The sequences given in R46793-824 are peptides derived from the
```

```
CC   phytase protein.  The phytase protein may be used in the composition
CC   of the invention.  The DNA encoding the phytase protein may be
CC   introduced into a Trichoderma host which then expresses it and the
CC   protein is collected from the culture medium.  By using Trichoderma as
CC   a host for Aspergillus phytate degrading enzymes such as this, a
CC   totally different enzyme composition compared to that secreted from
CC   Aspergillus results.  The enzyme composition can be used for removal
CC   of phytic acid or inositol hexaphosphoric acid from raw material,
CC   particularly plant material.  The composition is used in feed
CC   compositions for animals.  By using Trichoderma as a source of a
CC   composition containing phytate degrading enzymes some difficult
CC   downstream processing problems, eg. filtration, that occur with
CC   similar Aspergillus compositions are avoided and yields are improved.
XX
SQ   Sequence   7 AA;


  Query Match             42.2%;  Score 19;  DB 15;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+05;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 SGANLN 8
           | |:||
Db       1 saasln 6


Search completed: December 16, 2000, 03:07:17
Job time: 8058 sec
```